Image Reconstruction in Serial Femtosecond Nanocrystallography

Joe Chen

A thesis presented for the degree of
Doctor of Philosophy

in
Electrical and Computer Engineering
at the
University of Canterbury,
Christchurch, New Zealand.

July 2015

"The infinite we shall do right away. The finite may take a little longer."
Stanislaw Ulam

"All cells multiply by dividing."
Ian Stewart

"It's all Math!"
A disgruntled ENEL320 student

ABSTRACT

X-ray crystallography is a form of microscopy that allows the three-dimensional arrangement of atoms belonging to molecules within crystals to be determined. In this method, a crystal is illuminated with a beam of X-rays and the diffracted amplitudes resulting from the illumination are measured and computationally processed to enable the electron density of the unit molecule, or the unit cell, constituting the crystal to be calculated. The recent development of the X-ray free-electron laser (XFEL) provides new routes for determining molecular structures via its ability to generate intense but brief X-ray pulses. These new instruments enable diffraction measurements to be obtained from crystals that have a small number of unit cells, referred to as nanocrystals, and molecular structure determination via this technique is known as serial femtosecond nanocrystallography (SFX).

This thesis is concerned with the characterisation of diffraction data obtained from SFX experiments and the techniques for reconstructing the electron density of the molecule from such data. The noise characteristics of diffraction measurements from nanocrystals is developed. Methods for directly inverting nanocrystal diffraction to obtain the electron density of the molecule are analysed and an approach to ameliorate the effect of noise is proposed and evaluated by simulation. A model for diffraction by nanocrystals that include the effects of different unit cell arrangements and incomplete unit cells on the crystal surface is also developed and explored by simulation. The diffraction by finite crystals is shown to be equal to the incoherent average over a set of unit cells that contain different molecular arrangements related to the symmetry of the crystal at hand. The problem of image reconstruction under this circumstance is investigated. The more general problem of reconstructing multiple, unrelated, objects from their averaged diffraction is also explored and uniqueness properties along with reconstruction algorithms developed. The problem of reconstructing multiple, related, unit cells is studied and preliminary results are obtained. These results show that iterative phase retrieval algorithms can in principle be adapted to reconstruct the electron density of a crystalline specimen from the data obtained in SFX and the retrieval of phases from the diffracted intensity averaged over multiple objects is feasible.

CONTENTS

ACKNOWLEDGEMENTS

I am enormously grateful to my supervisors Rick Millane and Phil Bones, for all their time, guidance, stories and friendship. Their enthusiasm for what they do in academia and in life is truly inspirational.

During the course of my research I enjoyed a short visit to the Center for Free-Electron Laser Science (CFEL) in Hamburg, Germany, and on the same trip, participated in an experiment at the Linac Coherent Light Source (LCLS) near Stanford, CA, USA. I express my thanks to all the people I met during this trip, some of whom I don't even know by name. I especially thank Richard Bean who showed me around CFEL and provided support and guidance throughout my stay, Tom White for all the interesting discussions, Carolin Seuring for teaching me about the biological side of our research, and Rick Kirian for giving me valuable insights into the field of imaging with XFELs and a job at the end of my studies. I am also grateful to Irmtraud Kleine for taking care of the administration while I was in CFEL and to Henry Chapman for hosting my stay at the centre and driving the lot of us to the supermarket during a nightshift at the LCLS. I gratefully acknowledge the funds provided by the UC doctoral scholarship, the Canterbury branch of the Royal Society travel award, and the R. H. T. Bates scholarship which all helped in making this travel possible.

The other highlight in the course of my research was having the opportunity to teach an entire term of the 2nd professional year signal processing course in the Electrical and Computer Engineering department at UC. I am indebted to Phil Bones for giving me this opportunity and helping me throughout this great experience. I thank all the students who took part in ENEL320 in 2014. I have probably learnt more from you guys than you guys have from me.

I also thank my good friend and classmate Phil Postles, who shares with me the insatiable curiosity to understand the workings of our natural world. Finally, I would like to thank my mum and dad, who have always been there for me come rain, snow or shine.

PREFACE

I have always wanted to be an astronomer. Protein X-ray crystallography seems quite distinct from the field of astronomy. However, during the course of my research, I came to realise that X-ray crystallography is not too dissimilar to astronomy in many respects. In both disciplines a diffracted or propagated radiation field is measured, and the objects of interest, be they molecules or galaxies, are to be reconstructed from these measurements.

This thesis contains 8 chapters. Chapter 1 sets the scene and introduces the problem, Chapters 2-3 contain background material, and Chapters 4-7 report original work. A more detailed description of each chapter is as follows:

Chapter 2 reviews the mathematics and physics background associated with material presented in the later chapters, covering the basics of signal processing, probability, X-ray diffraction theory and iterative projection algorithms.

Chapter 3 outlines the problems associated with structure determination using conventional X-ray crystallography and then gives a tour of the method of serial femtosecond nanocrystallography that is able to solve these problems.

Chapter 4 presents a characterisation of the forward problem of diffraction by an ensemble of nanocrystals with the assumption that all unit cells are complete.

Chapter 5 explores the inverse problem of recovering the electron density when only the averaged magnitude of the Fourier transform from an ensemble of nanocrystals is available.

Chapter 6 characterises the forward problem of diffraction by an ensemble of nanocrystals with partial, or incomplete unit cells.

Chapter 7 is an account of the inverse problem of phase retrieval from the Fourier intensity averaged over multiple objects or multiple unit cells, a problem that arises due to the presence of incomplete unit cells described in Chapter 6.

Conclusions are drawn in Chapter 8 along with proposals for future work.

Publications that have resulted from the work presented in this thesis are listed below in reversed chronological order.

J. P. J. Chen, R. A. Kirian, K. R. Beyerlein, R. J. Bean, H. N. Chapman, P. J. Bones and R. P. Millane. "Phase retrieval for crystalline diffraction averaged over several different unit cells," in *Proceedings of the 2nd BioXFEL conference*, Puerto Rico, Jan 2015.

J. P. J. Chen, R. A. Kirian, K. R. Beyerlein, R. J. Bean, H. N. Chapman, P. J. Bones and R. P. Millane. "Phase retrieval from crystalline diffraction averaged over several different unit cells," in *Proceedings of the Image and Vision Computing New Zealand conference (IVCNZ)*, Hamilton, New Zealand, Nov 2014.

J. P. J. Chen and R. P. Millane."Diffraction by nanocrystals II," *Journal of the Optical Society of America A*, vol. 31, no. 8, pp. 1730-1737, 2014.

R. P. Millane and J. P. J. Chen."Aspects of direct phasing in femtosecond nanocrystallography," *Philosophical Transactions of the Royal Society B*, vol. 369, no. 20130498, June 2014.

K. R. Beyerlein, R. A. Kirian, R. J. Bean, H. N. Chapman, J. P. J. Chen and R. P. Millane. "Direct Phasing using Off-Bragg Intensity from Nanocrystals: Treatment of Different Crystal Edge Effects," in *Proceedings of the European XFEL Conference*, Hamburg, Germany, Jan 2014.

J. P. J. Chen, J. C. H. Spence and R. P. Millane. "Direct phasing in femtosecond nanocrystallography. II. Phase retrieval," *Acta Crystallographica A*, vol. 70, pp. 154-161, 2014.

J. P. J. Chen, J. C. H. Spence and R. P. Millane. "Direct phasing in femtosecond nanocrystallography. I. Diffraction characteristics," *Acta Crystallographica A*, vol. 70, pp. 143-153, 2014.

J. P. J. Chen and R. P. Millane."Diffraction by nanocrystals," *Journal of the Optical Society of America A*, vol. 30, no. 12, pp. 2627-2634, 2014.

J. P. J. Chen and R. P. Millane. "Diffraction by small crystals with incomplete unit cells," in *Proceedings of the Image and Vision Computing New Zealand conference (IVCNZ)*, Wellington, New Zealand, Nov 2013.

J. P. J. Chen, J. C. H. Spence and R. P. Millane "Aspects of Phasing Nanocrystal Diffraction Data From X-ray Free-electron Lasers," in *Proceedings of the Meeting onX-ray lasers in biology*, Chicheley, UK, Oct 2013.

J. P. J. Chen, R. P. Millane and J. C. H. Spence. "Phasing of selectively sampled data in nanocrystallography," in *Proceedings of the 11th International Conference on*

Biology and Synchrotron Radiation, Hamburg, Germany, Sep 2013.

J. P. J. Chen, J. C. H. Spence and R. P. Millane. "Phase retrieval in femtosecond X-ray nanocrystallography," in *Proceedings of the Image and Vision Computing New Zealand conference (IVCNZ)*, Dunedin, New Zealand, Nov 2012.

J. P. J. Chen, J. C. H. Spence and R. P. Millane. "Phase retrieval in nanocrystallography," in *Proceedings of SPIE: Image Reconstruction from Incomplete Data VII*, San Diego, CA, USA, Aug 2012.

J. P. J. Chen, J. C. H. Spence and R. P. Millane. "Aspects of phase retrieval for diffraction data from very small crystals," in *Proceedings of the 26th Image and Vision Computing New Zealand conference (IVCNZ)*, Auckland, New Zealand, Nov 2011.

J. P. J. Chen, J. C. H. Spence and R. P. Millane. "Phase retrieval in nanocrystallography," in *Proceedings of the ARC Centre of Excellence for Coherent X-ray Science 6th Annual Workshop*, Melbourne, Australia, Oct 2011.

Chapter 1

INTRODUCTION

The wheel of revolution began to spin in the early 20^{th} century when it was realised that the marriage of the newly discovered X radiation with the well-ordered composition of crystals gave rise to information that could be decoded to reveal the structure of the crystalline material. Thus was born the field of X-ray crystallography [Friedrich et al. 1912]. One hundred years later, the wheel of revolution has been set in motion once again with the arrival of a new generation of X-ray sources called the X-ray free-electron laser (XFEL). These large machines enable X-rays with unprecedented intensities to be achieved, packaged within coherent, and extraordinarily short pulses - shorter than even the time required for electronic transition processes in atoms [Chapman et al. 2014]. These three properties of X-rays: high intensity, full coherence and short pulse duration, have the potential to solve almost all major obstacles associated with the existing procedure of X-ray crystallography.

In crystallography, a crystalline material is probed with an external radiating source. X-rays are ideal atomic probes as their wavelengths are comparable to the linear scale of an atom, on the order of 10^{-10}m, designated the special unit of 1 Ångström (1Å). The penetrating power of X-rays also allows the three-dimensional structure of the material to be explored without being limited to the surface of the specimen. Any material that forms crystals can be subjected to this method of imaging. An important class of imaging targets exists in the field of structural biology, which is that of protein molecules. These large, so-called macromolecules are crucial in almost all biological processes occurring inside living cells with their functions being largely dictated by their three-dimensional structure [Lehninger et al. 2013]. Therefore, understanding the structures of proteins is important, for example, to be able to develop effective drugs for the treatment of diseases, to emulate natural processes such as photosynthesis for potential future clean energy generation, and not least to be able to understand life itself.

Conventional X-ray crystallography relies on the ready crystallisation of the molecule of interest. The crystallised product is a periodic structure composed of identical, repeating molecules of the specimen under study, with a period of this periodic structure

being labelled as a *unit cell*. An X-ray beam is then made to collide with the crystal and the interference of the X-rays scattering off the electrons of the atoms inside the crystal generates a *diffraction pattern* which is recorded. From this diffraction pattern, corresponding to the squared magnitude of the Fourier transform of the electron distribution inside the crystal, a map of the electron density can be deduced by utilising various methods to recover the Fourier phase. The lack of phase information from X-ray diffraction experiments constitutes the so-called *phase problem*. Upon solving the phase problem and deducing the electron density map, positions of individual atoms can in turn be inferred. Thus in effect, X-ray crystallography provides an atomic resolution image of the molecule.

The whole process runs aground however if the sample of interest cannot be readily crystallised. This simple limitation spurred much of the effort to extend traditional crystallography techniques to crystals with a smaller number of unit cells, termed *nanocrystals*, and even to non-crystalline samples in general, i.e. single molecules. Such an extension is not trivial however. The most obvious obstacle is that for crystals of small sizes, the intensities of the diffracted X-rays are nowhere near the detection levels necessary for current detectors to observe their diffraction pattern. The registered diffraction signal essentially becomes engulfed by noise for crystals with a very small number of unit cells. The most straightforward way to overcome this problem is simply to increase the intensity of the incident X-ray beam. This results in a higher photon count on the detector and gives a better signal-to-noise ratio (SNR). However, the high X-ray intensity causes the biological specimen to sustain severe radiation damage which may result in a change in its intrinsic structure, or even worse - its destruction. A solution to this conundrum is provided by the newly emerging technology of free-electron lasers in the X-ray region of the electromagnetic spectrum.

X-ray free-electron lasers are devices integrated within particle accelerators in order to achieve their characteristic high intensity, short duration X-ray pulses. The duration of these pulses can currently be made to be on the order of 10^{-15} seconds, or femtoseconds (fs). Because the electron transfer dynamics of molecules and atoms are also on timescales on the orders of femtoseconds, information about basic reaction mechanisms such as chemical bond forming/breaking could be obtained by probing structures with these femtosecond excitations. Furthermore, such XFEL pulses are so brief that diffraction patterns can be recorded before radiation damage is able to manifest itself on the sample sustaining bombardment, thereby enabling the possibility of capturing an image of the sample before it is vaporised by the intense X-ray pulse. Meaningful diffraction signal from small crystals can therefore be captured using XFELs.

Being able to use smaller crystals not only avoids the need for tedious crystallisation, but it also potentially allows the phases of the diffracted field to be recovered more easily by providing a glimpse of the continuous diffraction pattern from a single unit cell. This is possible as crystals with a small enough number of unit cells give rise to

measurable diffraction intensities that can be sampled more finely than for crystals with a large number of unit cells. Although these additional samples are highly attenuated, they offer the possibility of phase retrieval without the use of ancillary experimental data. Together with XFELs, the use of small crystals has given rise to a new variant of X-ray crystallography called *serial femtosecond nanocrystallography* (SFX).

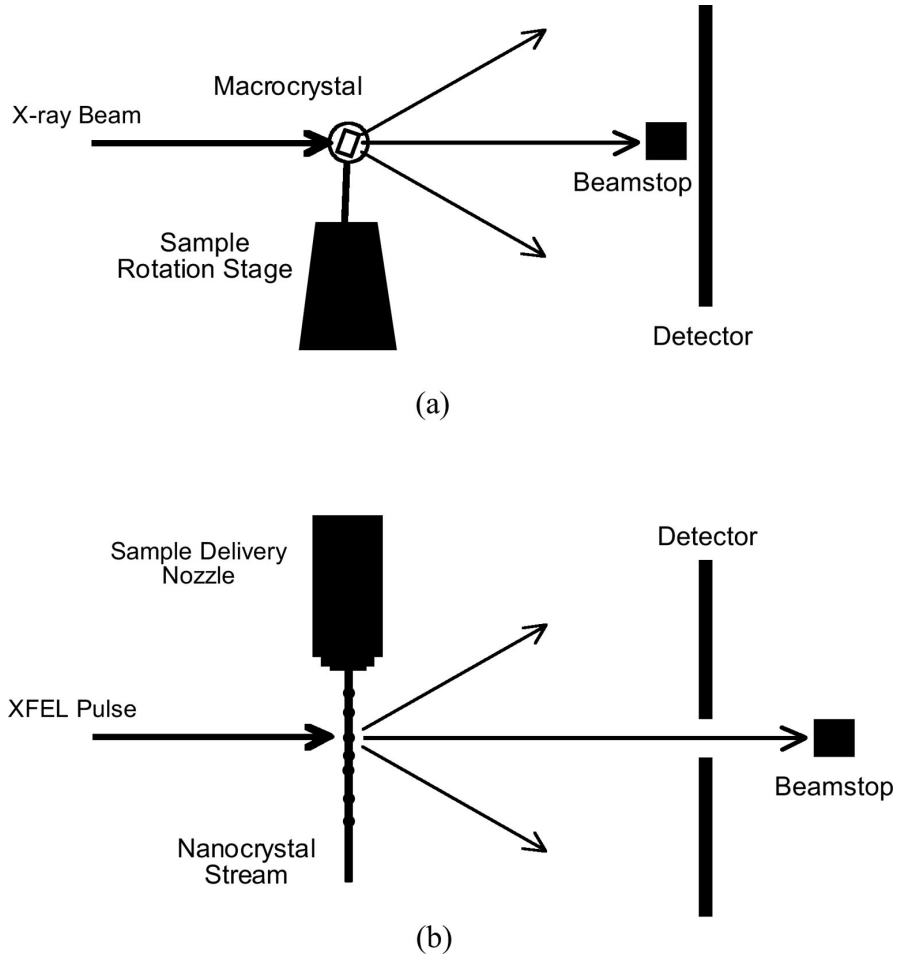

Figure 1.1 Schematic diagrams of the experimental setup for (a) conventional X-ray crystallography and (b) serial femtosecond nanocrystallography.

Differences between the experimental setup in conventional X-ray crystallography and that of SFX can be seen in Fig. 1.1. Because the intensity of the XFEL pulse is extremely high, the beamstop can no longer be placed in front of the detector but a gap must be designed on the detector for the beam to pass through and de-focus before being stopped. Moreover, for the same reason, the sample will be destroyed after being hit by a single XFEL pulse. The destruction of the sample shortly after its exposure to the X-rays means that a new sample must be moved into the path of the X-ray beam in time for the next pulse. One method by which this can be achieved is to deliver the samples to the X-ray interaction region as a liquid suspension using a suitable injection device [DePonte et al. 2008, Weierstall et al. 2012], as shown in Fig. 1.1(b).

The consequence of this delivery method is that a large number of diffraction patterns from crystals of unknown sizes, shapes and orientations are collected within the course of the SFX experiment. The challenge in processing the data from SFX then involves reducing the raw data volume obtained in experiments to manageable amounts, and analysing the resulting reduced data to extract the structure of the molecule from the combined and averaged diffraction patterns.

Huge advances have already been made on the interpretation and conversion of the data into an image of the molecular structure from which the data was derived [Spence et al. 2012]. It turns out that diffraction patterns collected from an SFX experiment can be processed using modified versions of methods already in use in conventional X-ray crystallography [Kirian et al. 2010, Kirian et al. 2011], with the main difference being that a large number of patterns must be processed and the resulting intensities combined in a Monte Carlo fashion. The merging of the diffraction data in this way is one of many new methods and ideas that the emerging field of SFX has prompted researchers to pursue. Another avenue that has arisen from the SFX approach is a novel method of solving the phase problem by accessing the finely sampled diffraction intensities from small crystals as mentioned above [Spence et al. 2011]. This new idea is tested and explored further in this thesis.

For large crystals, properties such as the configuration of their constituent molecules on their surface have little effect on the resulting diffraction. For nanocrystals on the other hand, these effects can no longer be ignored and are instead key clues in understanding the SFX method and for it to be properly utilised in structure determination. The effect of surface structures of crystals on their diffraction is investigated in the second part of this thesis. The work presented here therefore contributes to an understanding of the paradigm shift from large, macroscopic crystals to nanocrystals with a small number of unit cells.

The method of SFX, and in general, the imaging of matter with XFELs, is still in its infancy. But given enough time and effort, it should mature into a technique able to provide valuable insights into the workings of the molecular world.

Chapter 2

MATHEMATICS AND PHYSICS BACKGROUND

The relevant mathematical and physical understandings underpinning much of the material in this thesis are presented in this chapter, with the aim of establishing notations and preparing the reader for the discussions in the chapters to follow.

2.1 SIGNALS, IMAGES AND FUNCTIONS

Reality can be characterised by numbers and the relationships between numbers are therefore of great interest. Mappings are prescribed between one set of numbers and another, usually one-to-one, where one set of numbers could be time instants for a temporal signal, or coordinates of points within an n-dimensional (n-D) space. The second corresponding set of numbers is then the value of the temporal signal/spatial image at those times/locations. The set of numbers $\mathbf{x}$ together with a mapping f defines a particular "signal," "image" or "function." These terms are treated synonymously in this thesis.

2.1.1 Properties of Signals

A signal $f(\mathbf{x})$ is said to be periodic with period $\mathbf{T}$ if there exists a $\mathbf{T} \neq 0$ such that

$$f(\mathbf{x}) = f(\mathbf{x} + \mathbf{T}). \tag{2.1}$$

The $\mathbf{T}$ with the smallest magnitude for which Eq. (2.1) holds is called the fundamental period of the signal $f(\mathbf{x})$. Every integer multiple of the fundamental period is also a period of the signal. A signal is bounded above and below if there exists an A such that $|f(\mathbf{x})| < A < \infty$, and the signal is of finite extent, or compactly supported, if $f(\mathbf{x})$ is zero for all $\mathbf{x}$ outside some closed set S. The set S is then known as the *support* of the signal. A signal is *discrete* when the variable $\mathbf{x}$ can be enumerated by the integers.

Knowing that a particular signal has properties such as those described above gives rise to powerful prior information about the signal which can be used to narrow the search space should the signal need to be recovered.

2.1.2 The Fourier Transform

One way to analyse a complicated signal is to decompose it into simpler constituent signals. One such class of simple signals are sinusoids. The mathematics of Fourier analysis encapsulates the decomposition of a signal into a sum of sinusoids of different amplitudes, frequencies and phases. The Fourier transform is therefore able to provide an alternative, often simpler, view of a signal. This fact, coupled with the existence of fast algorithms for carrying out Fourier transformations on a digital computer, gives us an effective way to not only describe, but to also measure and simulate reality.

Let $\mathbf{x} = [x_1, \ldots, x_n]^T$ and $\mathbf{u} = [u_1, \ldots, u_n]^T$ denote continuous positions in an n-D spatial and spatial frequency domain, respectively, such that x_j and u_j are real numbers for all $j = 1, \cdots, n$. The Fourier transform $F(\mathbf{u})$ of a continuous function $f(\mathbf{x})$ is defined as

$$F(\mathbf{u}) = \int_{-\infty}^{\infty} f(\mathbf{x}) \exp\left(-i2\pi\mathbf{u}\cdot\mathbf{x}\right) \mathrm{d}\mathbf{x}, \tag{2.2}$$

and the inverse Fourier transform is given by

$$f(\mathbf{x}) = \int_{-\infty}^{\infty} F(\mathbf{u}) \exp\left(i2\pi\mathbf{x}\cdot\mathbf{u}\right) \mathrm{d}\mathbf{u}, \tag{2.3}$$

where it is understood that the boldface differential operators $\mathrm{d}\mathbf{x}$ and $\mathrm{d}\mathbf{u}$ mean that integration is over all components of the vectors $\mathbf{x}$ and $\mathbf{u}$. In the field of X-ray crystallography, the domains indexed by $\mathbf{x}$ and $\mathbf{u}$ are referred to respectively as the *real space* and the *reciprocal space*. The signal is therefore seen from two different perspectives through the Fourier transformation. As a shorthand, the Fourier transform and its inverse will also be written in this thesis as $F(\mathbf{u}) = \mathcal{F}\{f(\mathbf{x})\}$ and $f(\mathbf{x}) = \mathcal{F}^{-1}\{F(\mathbf{u})\}$, respectively.

A function is guaranteed to have a Fourier transform if it fulfils the Dirichlet conditions:

1. $f(\mathbf{x})$ is square-integrable, i.e. $\int_{-\infty}^{\infty} |f(\mathbf{x})|^2 \, \mathrm{d}\mathbf{x}$ is finite.
2. $f(\mathbf{x})$ is piece-wise continuous.
3. $f(\mathbf{x})$ is bounded above and below.
4. $f(\mathbf{x})$ is single-valued.

A function is piecewise continuous if it is continuous on all but a finite number of points. However, the Dirichlet conditions are *sufficient* in that the Fourier transform of a function exists if they are satisfied; they do not say anything about the transformability for functions that violate one or more of these conditions. Thus the Fourier transform of a function could still exist even if that function does not obey all of the Dirichlet criteria.

An important example of such a function is the delta function of Dirac, defined loosely as

$$\delta(\mathbf{x}) = \begin{cases} \infty & \mathbf{x} = \mathbf{0} \\ 0 & \mathbf{x} \neq \mathbf{0} \end{cases} \tag{2.4}$$

with the requirement that

$$\int_{0^-}^{0^+} \delta(\mathbf{x}) \mathrm{d}\mathbf{x} = 1. \tag{2.5}$$

The delta function has the property

$$\int_{-\infty}^{\infty} \delta(\mathbf{x} - \boldsymbol{\xi}) f(\mathbf{x}) \mathrm{d}\mathbf{x} = f(\boldsymbol{\xi}), \tag{2.6}$$

for any continuous function $f(\mathbf{x})$. No function in the ordinary sense has the property given by Eq. (2.6) but if we define a sequence of functions which progressively grows taller and thinner at $\mathbf{x} = \boldsymbol{\xi}$, with the area underneath each member function of the sequence remaining equal to 1, we can see that the value of the limit function tends to 0 at every point except at $\mathbf{x} = \boldsymbol{\xi}$ where it tends to infinity. The limit function of this sequence, referred to as a *generalised function*, would then have the property given by Eq. (2.6), called the *sifting property of the Dirac delta*. On application of the sifting property, it can be seen that $\mathcal{F}\{\delta(\mathbf{x} - \boldsymbol{\xi})\} = \exp(-i2\pi(\mathbf{u} \cdot \boldsymbol{\xi}))$. The delta function is thus a prime example of a signal that does not satisfy the Dirichlet conditions (by not being bounded) but still possesses a Fourier transform.

Another class of functions where the Dirichlet criteria do not directly apply but still has Fourier transforms is that of discrete functions. Let $\mathbf{x}_\mathrm{D} = [x_1, \ldots, x_n]^T$ and $\mathbf{u}_\mathrm{D} = [u_1, \ldots, u_n]^T$ be coordinate vectors of an n-D discrete spatial and spatial frequency domain, respectively, such that x_j and $u_j \in \{0, \cdots, N_j - 1\}$ for all $j = 1, \cdots, n$ where N_j is a whole number denoting the length of the discrete signal in the jth dimension. An n-D discrete signal $f(\mathbf{x}_\mathrm{D})$ then has the discrete Fourier transform (DFT) $F(\mathbf{u}_\mathrm{D})$ given by

$$F(\mathbf{u}_\mathrm{D}) = \sum_{\mathbf{x}_\mathrm{D}=0}^{(\mathrm{N}-\mathrm{I})_{ii}} f(\mathbf{x}_\mathrm{D}) \exp\left(-i2\pi \mathbf{u}_\mathrm{D}^T \, \mathrm{N}^{-1} \mathbf{x}_\mathrm{D}\right), \tag{2.7}$$

and the inverse discrete Fourier transform,

$$f(\mathbf{x}_\mathrm{D}) = \frac{1}{\det(\mathrm{N})} \sum_{\mathbf{u}_\mathrm{D}=0}^{(\mathrm{N}-\mathrm{I})_{ii}} F(\mathbf{u}_\mathrm{D}) \exp\left(i2\pi \mathbf{x}_\mathrm{D}^T \, \mathrm{N}^{-1} \mathbf{u}_\mathrm{D}\right). \tag{2.8}$$

The symbol $\det(\cdot)$ denotes the determinant of a square matrix, $(\cdot)_{ii}$ represents the

matrix diagonal elements, and

$$\mathrm{N} = \begin{bmatrix} N_1 & 0 & \cdots & 0 \\ 0 & N_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & N_n \end{bmatrix}. \tag{2.9}$$

The vectorised summation $\sum_{\mathbf{x}_\mathrm{D}=0}^{(\mathrm{N}-\mathrm{I})_{ii}}$ is understood to be a shorthand for $\sum_{x_n=0}^{N_n-1} \cdots \sum_{x_1=0}^{N_1-1}$.

An important property of the Fourier transform is that $f(\mathbf{x})$ and $F(\mathbf{u})$ are in general complex-valued functions. They are in turn expressible as a sum of their real and imaginary parts, denoted in this thesis by $Re(\cdot)$ and $Im(\cdot)$. Furthermore, they are also expressible as a product of their magnitude and complex exponentiated phase components, denoted by $|\cdot|$ and ϕ, respectively. $F(\mathbf{u})$, for example, can be written as

$$F(\mathbf{u}) = Re\left(F(\mathbf{u})\right) + Im\left(F(\mathbf{u})\right) = |F(\mathbf{u})| \exp(i\phi(\mathbf{u})). \tag{2.10}$$

If $f(\mathbf{x})$ is real, i.e., $Im(f(\mathbf{x})) = 0$, then its Fourier transform is Hermitian, satisfying

$$F(-\mathbf{u}) = F^*(\mathbf{u}) \tag{2.11}$$

where the asterisk denotes complex conjugation. In crystallography, the Hermitian relationship is known as Friedel's Law and the centrosymmetric points $\mathbf{u}$ and $-\mathbf{u}$ are called Friedel pairs.

2.1.3 Convolution and Multiplication

The convolution of two continuous signals, $f(\mathbf{x})$ and $g(\mathbf{x})$, is defined by

$$f(\mathbf{x}) \otimes g(\mathbf{x}) = \int_{-\infty}^{\infty} f(\boldsymbol{\xi}) g(\mathbf{x} - \boldsymbol{\xi}) \mathrm{d}\boldsymbol{\xi}. \tag{2.12}$$

The correlation of two continuous signals, $f(\mathbf{x})$ and $g(\mathbf{x})$, can then be written as

$$f(\mathbf{x}) \odot g(\mathbf{x}) = \int_{-\infty}^{\infty} f^*(\boldsymbol{\xi}) g(\mathbf{x} + \boldsymbol{\xi}) \mathrm{d}\boldsymbol{\xi} = f^*(-\mathbf{x}) \otimes g(\mathbf{x}). \tag{2.13}$$

The convolution operation is associative, distributive and commutative. The convolution theorem states that the operation of multiplication in one domain becomes that of convolution in the Fourier transform of that domain, i.e.

$$\mathcal{F}\{f(\mathbf{x}) \otimes g(\mathbf{x})\} = F(\mathbf{u})G(\mathbf{u}) \tag{2.14}$$

and vice versa, making Fourier transforms useful in turning difficult-to-solve integral problems into potentially easier-to-solve algebraic problems.

It is straightforward to show that the Fourier transform of the *autocorrelation* of a signal is equal to the squared magnitude of the Fourier transform of that signal, i.e.

$$\mathcal{F}\{f(\mathbf{x}) \otimes f^*(-\mathbf{x})\} = |F(\mathbf{u})|^2 . \tag{2.15}$$

Setting $\mathbf{x} = \mathbf{0}$ and taking the inverse Fourier transform of both sides in Eq. (2.15) leads to Parseval's theorem which is a statement of the preservation of Euclidean distance (or energy) by the Fourier transform,

$$\int_{-\infty}^{\infty} |f(\mathbf{x})|^2 \, \mathrm{d}\mathbf{x} = \int_{-\infty}^{\infty} |F(\mathbf{u})|^2 \, \mathrm{d}\mathbf{u}. \tag{2.16}$$

Plancherel [1910] generalised this to two complex and distinct functions by his relation

$$\int_{-\infty}^{\infty} f(\mathbf{x}) \, g^*(\mathbf{x}) \, \mathrm{d}\mathbf{x} = \int_{-\infty}^{\infty} F(\mathbf{u}) \, G^*(\mathbf{u}) \, \mathrm{d}\mathbf{u}. \tag{2.17}$$

2.1.4 Useful Formalisms and Functions

Besides the delta function, other functions that are of use in this thesis include the rectangular function, rect($\mathbf{x}$), and the sinc function, sinc($\mathbf{x}$). The 1-D rectangular function is defined as

$$\mathrm{rect}(x) = \begin{cases} 1 & |x| \leq \frac{1}{2} \\ 0 & \text{otherwise.} \end{cases} \tag{2.18}$$

The sinc function is defined in 1-D as

$$\mathrm{sinc}(x) = \frac{\sin(\pi x)}{\pi x}. \tag{2.19}$$

In n-D, the two functions can be defined as a product of their 1-D counterparts, specifically,

$$\mathrm{rect}(\mathbf{x}) = \prod_{j=1}^{n} \mathrm{rect}(x_j), \tag{2.20}$$

and

$$\mathrm{sinc}(\mathbf{x}) = \prod_{j=1}^{n} \mathrm{sinc}(x_j). \tag{2.21}$$

The rect and the sinc functions form a Fourier transform pair.

The comb function is defined as an infinite summation of uniformly spaced delta functions. In 1-D, it is

$$\mathrm{comb}\left(\frac{x}{a}\right) = a \sum_{m=-\infty}^{\infty} \delta(x - am) \tag{2.22}$$

where the parameter a controls the spacings of the individual delta functions. In n-D,

the comb function can be defined as

$$\operatorname{comb}\left(\mathrm{L}^{-1}\mathbf{x}\right) = \det(\mathrm{L}) \sum_{\mathbf{m}=-\infty}^{\infty} \delta(\mathbf{x} - \mathrm{L}\mathbf{m}), \tag{2.23}$$

where $\mathbf{m}$ is an n-tuple of integers and L is a matrix defined for a rectangular lattice of delta functions with a constant spacing a_j in the jth dimension as

$$\mathrm{L} = \begin{bmatrix} a_1 & 0 & \cdots & 0 \\ 0 & a_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & a_n \end{bmatrix}. \tag{2.24}$$

L is referred to here as the lattice matrix.

The Fourier transform of a comb function is another comb function but with a lattice matrix equal to the inverse of that for the original comb in the other domain, i.e.

$$\mathcal{F}\left\{\operatorname{comb}\left(\mathrm{L}^{-1}\mathbf{x}\right)\right\} = \det(\mathrm{L})\,\operatorname{comb}(\mathrm{L}\mathbf{u}), \tag{2.25}$$

which is generalised for any Fourier transformable function f by the Poisson summation formula [Blahut 2011]

$$\sum_{\mathbf{m}=-\infty}^{\infty} f(\mathrm{L}^{-1}\mathbf{m}) = \sum_{\mathbf{h}=-\infty}^{\infty} F(\mathrm{L}\mathbf{h}). \tag{2.26}$$

Putting to use the convolution theorem from the previous section, Eq. (2.25) says that an infinitely repeating signal (described by the convolution with a comb function) has a discrete spectrum (described by the multiplication with a comb function). Using the definition of the comb function given by Eq. (2.22), Eq. (2.25) can be written explicitly in 1-D as

$$\mathcal{F}\left\{\sum_{m=-\infty}^{\infty} \delta(x - am)\right\} = \frac{1}{a} \sum_{h=-\infty}^{\infty} \delta\left(u - \frac{h}{a}\right). \tag{2.27}$$

In general, L is a non-singular n-by-n matrix that defines a lattice in n-D space such that the set of all points on that lattice have coordinates given by

$$\mathbf{x}_{\mathbf{m}} = \mathrm{L}\mathbf{m}, \tag{2.28}$$

and the corresponding lattice in Fourier space is

$$\mathbf{u}_{\mathbf{h}} = \mathrm{L}^{-1}\mathbf{h}. \tag{2.29}$$

The column vectors $\mathbf{m}$ and $\mathbf{h}$ are both an n-tuple of integers. This formalism can also be used to describe the electron density of a crystal whereby a 3-by-3 lattice matrix is defined that specifies the repetition of a unit cell throughout 3-D space. In this case,

$\mathbf{x_m}$ specifies the points on the so-called *real space lattice* and $\mathbf{u_h}$ represent points on the *reciprocal lattice*. As shall be seen in Sec. 2.3.4, the lattice points given by the reciprocal lattice are the positions of the so-called *Bragg peaks* in a crystallographic diffraction pattern.

2.1.5 Sampling and Interpolation

To convert a continuous signal into a discrete form, it needs to be sampled. As alluded to in the previous section, the sampling process can be described as the multiplication of the signal to be sampled with a comb function. Illustrating this first in 1-D, sampling a continuous signal $f(x)$ to give a discrete signal $f(x_{\mathrm{D}})$ can be written as

$$f(x_{\mathrm{D}}) = \mathrm{comb}\left(\frac{x}{a}\right) f(\mathbf{x}). \tag{2.30}$$

Provided that the samples are spaced finely enough, the relation that justifies the uniqueness between the original continuous signal and its equidistant samples is the Whittaker-Kotel'nikov-Shannon sampling theorem, commonly known simply as the Shannon sampling theorem [Shannon 1949]. The theorem says that if a signal $f(x)$ has a Fourier transform $F(u)$ that is zero everywhere except within a finite interval $\mathcal{S}$, i.e. the support of $F(u)$ is finite, then the largest spacing of the samples of $f(x)$ that can still completely determine the signal is $1/|\mathcal{S}|$. The operator $|\cdot|$ in this case denote the length of the interval. The original continuous signal $f(x)$ can be reconstructed from its samples $f(x_{\mathrm{D}})$ via the equation

$$f(x) = \sum_{h=-\infty}^{\infty} f\left(\frac{h}{|\mathcal{S}|}\right) \frac{\sin\left(\pi|\mathcal{S}|\left(x - h/|\mathcal{S}|\right)\right)}{\pi|\mathcal{S}|\left(x - h/|\mathcal{S}|\right)} = f(x_{\mathrm{D}}) \otimes \mathrm{sinc}(|\mathcal{S}|x). \tag{2.31}$$

The above process is an interpolation and the sinc function is the interpolating function in this case. The domain of sampling and finite support are interchangable. This relation will be referred to in this thesis as "the sampling theorem." For an interesting review on the historical developments of this theorem, see Zayed [1993]. The sampling frequency, $|\mathcal{S}|$ inverse units, needed for the complete determination of the continuous signal from its samples is referred to as the Nyquist sampling rate in the engineering literature.[1]

In n-D, sampling a signal $f(\mathbf{x})$ to give $f(\mathbf{x}_{\mathrm{D}})$ can be described as the multiplication

[1] In the field of communications engineering the sampling theorem is often stated as "If a signal $f(x)$ contains no frequency components greater than B cycles per unit, then $f(x)$ is completely determined by its values at a sequence of locations spaced $1/(2B)$ units apart." The reason for the extra factor of 2 here is due to the bandwidth B of a signal in communications being defined as $|S|/2$. Since signals in this situation are real, their Fourier transforms are symmetric about the origin, and thus only half of the spectrum is needed to define it.

of $f(\mathbf{x})$ by an n-D comb function, i.e.

$$f(\mathbf{x}_{\mathrm{D}}) = \mathrm{comb}\left(\mathrm{L}^{-1}\mathbf{x}\right) f(\mathbf{x}), \tag{2.32}$$

and reconstruction can be achieved under ideal conditions by convolving the samples with an n-D sinc function.

$$f(\mathbf{x}) = \mathrm{sinc}(\mathrm{L}\mathbf{x}) \otimes f(\mathbf{x}_{\mathrm{D}}). \tag{2.33}$$

Other types of interpolating functions are possible and sampling on a lattice other than a rectangular grid can also be done.

2.2 PROBABILITY AND STATISTICS

The most convenient, and often the only tractable way, of analysing systems that contain a large number of parameters and/or vary in some random fashion is through probabilistic and statistical methods. The variability in the outcomes of any physical experiment is one example that requires such a formalism.

2.2.1 Probability Density and Cumulative Function

A random variable is a variable whose value depends on the outcome of a random event. These outcomes may be discrete, such as the two sides of a coin, or they may be continuous, such as the velocity of particles in a gas. Capital letters shall be used to denote a random variable and lower case letters represent a value that it takes. For any random variable X, each outcome x is assigned a non-negative value between 0 and 1 inclusive that describes the likelihood of that random variable taking on that particular outcome. This non-negative value is denoted $Prob(X = x)$ and is the probability of X being x. The probabilities of all possible outcomes add to one. The material in this section is presented in the context of continuous random variables but their discrete form can be obtained by replacing the integral operators with summations.

The cumulative distribution function (CDF) of a random variable X, denoted by $P_X(x)$, is the probability that X is less than or equal to a specific value x, which can be written as

$$P_X(x) = Prob(X \leq x). \tag{2.34}$$

The probability density function (PDF) of X, denoted by $p_X(x)$, is then defined as

$$p_X(x) = \frac{\mathrm{d}}{\mathrm{d}x} P_X(x), \tag{2.35}$$

and X is said to be "distributed as" $p_X(x)$, denoted by $X \sim p_X$. The probability of

the random variable X taking on a value on the interval $[a, b]$ can be found by

$$Prob(a \leq X \leq b) = \int_a^b p_X(x)\mathrm{d}x. \tag{2.36}$$

For more than one random variable acting together at the same time, a *joint random variable* can be defined. Using just two random variables to illustrate, the joint random variable XY is defined to be the collection of all possible joint outcomes (x, y) together with an associated measure of probability. The CDF for the joint random variable XY is defined as

$$P_{XY}(x, y) = Prob(X \leq x \ \cap \ Y \leq y), \tag{2.37}$$

where $\cap$ denotes intersection of events. The joint PDF is then

$$p_{XY}(x, y) = \frac{\partial^2}{\partial x \partial y} P_{XY}(x, y), \tag{2.38}$$

and the probability that X lies within the interval $[a, b]$ and Y within $[c, d]$ can be found by

$$Prob(a \leq X \leq b \ \cap \ c \leq Y \leq d) = \int_a^b \int_c^d p_{XY}(x, y)\mathrm{d}y\mathrm{d}x. \tag{2.39}$$

If we wish to determine the probability that an event described by the random variable X occurs regardless of the outcome of a particular event governed by the random variable Y, then we can specify a *marginal* probability density function. Marginal PDFs of the joint random variable XY are given by

$$p_X(x) = \int_{-\infty}^{\infty} p_{XY}(x, y)\mathrm{d}y, \tag{2.40}$$

and

$$p_Y(y) = \int_{-\infty}^{\infty} p_{XY}(x, y)\mathrm{d}x. \tag{2.41}$$

The probability that an event described by the random variable X occurs given that the outcome of another event governed by the random variable Y is known is referred to as the *conditional* probability, written as $Prob(X|Y)$. The conditional PDFs of the joint random variable XY are defined as

$$p_{Y|X}(y|x) = \frac{p_{XY}(x, y)}{p_X(x)}, \tag{2.42}$$

and

$$p_{X|Y}(x|y) = \frac{p_{XY}(x, y)}{p_Y(y)}. \tag{2.43}$$

Two random variables X and Y are *statistically independent* if knowledge about the outcome of one does not influence the probabilities of the outcomes of the other. By Eqs. (2.42) and (2.43), this implies that $p_{Y|X}(y|x) = p_Y(y)$, $p_{X|Y}(x|y) = p_X(x)$ and

the joint PDF of two statistically independent random variables is separable, i.e.

$$p_{XY}(x,y) = p_X(x)p_Y(y). \tag{2.44}$$

2.2.2 Statistical Averages

If x is the outcome of a random variable X with PDF $p_X(x)$, and g is a real function, then the statistical average, or the expected value, of $g(X)$ is defined by

$$E\left(g(X)\right) = \int_{-\infty}^{\infty} g(x)p_X(x)\mathrm{d}x. \tag{2.45}$$

The nth moment of a continuous PDF about a constant α can be found by setting $g(X) = (X-\alpha)^n$, giving

$$E\big(\,(X-\alpha)^n\,\big) = \int_{-\infty}^{\infty} (x-\alpha)^n\, p_X(x)\mathrm{d}x. \tag{2.46}$$

The *mean*, denoted by $\bar{x}$ or $E\left(X\right)$, is defined as the first moment centred about $\alpha = 0$,

$$E\left(X\right) = \bar{x} = \int_{-\infty}^{\infty} xp_X(x)\mathrm{d}x. \tag{2.47}$$

The *variance*, denoted by σ^2 or $Var\left(X\right)$, is the second moment centred about $\alpha = \bar{x}$,

$$Var\left(X\right) = E\big(\,(X-\bar{x})^2\,\big) = \sigma^2 = \int_{-\infty}^{\infty} (x-\bar{x})^2\, p_X(x)\mathrm{d}x. \tag{2.48}$$

The square root of the variance is the *standard deviation* σ and is a measure of the spread of the values assumed by the random variable X.

Expectation and variance obey the following rules

$$E\left(aX+bY+c\right) = aE\left(X\right) + bE\left(Y\right) + c, \tag{2.49}$$

and

$$Var\left(aX+bY+c\right) = a^2 Var\left(X\right) + b^2 Var\left(Y\right), \tag{2.50}$$

where a, b, c are constants and the random variables X and Y are statistically independent.

2.2.3 Correlations

When two random variables exist, their (n,m)th joint moment is defined as

$$E\big(\,(X-\alpha)^n\,(Y-\beta)^m\,\big) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x-\alpha)^n\,(y-\beta)^m\, p_{XY}(x,y)\mathrm{d}x\mathrm{d}y. \tag{2.51}$$

Of particular interest are the cases when $n = m = 1$, $\alpha = \bar{x}$ and $\beta = \bar{y}$, called the *covariance* of X and Y, and when $n = m = 1$, $\alpha = 0$ and $\beta = 0$ called the *correlation* of X and Y. The covariance is written as $Cov(X, Y)$ or σ^2_{XY}. The correlation coefficient, ρ, sometimes also referred to as Pearson's correlation coefficient, is then defined by

$$\rho = \frac{E\big(\,(X - \bar{x})\,(Y - \bar{y})\,\big)}{\sigma_X \sigma_Y} = \frac{Cov(X, Y)}{\sigma_X \sigma_Y}. \tag{2.52}$$

The correlation coefficient always has a value within ± 1 and is a measure of the similarity between X and Y, in particular, it measures how close X and Y are to being linearly related. When $\rho = 0$, X and Y are said to be uncorrelated. Two statistically independent random variables are always uncorrelated, however two uncorrelated random variables are not necessarily statistically independent. One exception when uncorrelatedness does imply independence is when the two random variables are *jointly* normally distributed, as to be defined next.

2.2.4 Important Probability Distributions

Several PDFs that will come into play later in this thesis and are important in their own right are discussed here. The continuous versions of the PDFs are shown, but their discrete versions can be obtained by sampling the continuous PDF and then normalising the samples such that the total probabilities sum to one.

2.2.4.1 Uniform Distribution

A random variable X is uniformly distributed on the interval $[a, b]$ if it has the PDF

$$p_X(x) = \begin{cases} \frac{1}{b-a} & a \leq x \leq b \\ 0 & \text{otherwise.} \end{cases} \tag{2.53}$$

It has a mean $E(X) = (a + b)/2$ and a variance $Var(X) = (1/12)(a - b)^2$.

2.2.4.2 Normal Distribution

A collection of measurements in nature frequently obeys a normal, or Gaussian, distribution, characterised by a symmetric bell-shaped curve. A random variable X is normal if it has the PDF

$$p_X(x) = \mathcal{N}\left(\bar{x}, \sigma^2\right) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x - \bar{x})^2}{2\sigma^2}\right). \tag{2.54}$$

Writing n jointly Gaussian random variables as the vector $\mathbf{X} = [X_1, \cdots, X_n]^T$ and denote the values that these random variables can assume by $\mathbf{x} = [x_1, \cdots, x_n]^T$, with

another vector containing their respective means denoted by $\bar{\mathbf{x}} = [\bar{x}_1, \cdots, \bar{x}_n]^T$, the joint PDF of these n jointly Gaussian random variables can be written as

$$p_{\mathbf{X}}(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^n \det(\mathrm{C})}} \exp\left(-\frac{1}{2} (\mathbf{x} - \bar{\mathbf{x}})^T \mathrm{C}^{-1} (\mathbf{x} - \bar{\mathbf{x}}) \right), \tag{2.55}$$

where C is an n-by-n matrix called the covariance matrix with the entry in the ith row and jth column defined by

$$\sigma_{ij}^2 = E\big((X_i - \bar{x}_i)(X_j - \bar{x}_j) \big). \tag{2.56}$$

The normal distribution is important as it describes the distribution of the sum of a large number of random variables, a result referred to as the *central limit theorem*. More specifically, let $X_1, \cdots, X_n$ be n independent random variables each with an arbitrary PDF that has means $\bar{x}_1, \cdots, \bar{x}_n$ and variances $\sigma_1^2, \cdots, \sigma_n^2$, respectively. Introduce a random variable Z defined by

$$Z = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{X_i - \bar{x}_i}{\sigma_i}. \tag{2.57}$$

Then under certain conditions, to be stated below, the PDF $p_Z(z)$ approaches a normal distribution with a mean of 0 and a standard deviation of 1 as the total number of variables n tends to infinity, i.e.,

$$\lim_{n\to\infty} p_Z(z) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2} \right). \tag{2.58}$$

This is the central limit theorem. While many variants of necessary and sufficient conditions exist for this theorem (see Goodman [1985] and references therein), of particular interest to us is the situation when the random variables X_i are independent but not identically distributed. A set of conditions due to Lyapunov states that if the mean and variance of all the X_is are finite and the condition

$$\lim_{n\to\infty} \frac{\sum_{i=1}^{n} E\left(|X_i - \bar{x}_i|^{2+\beta} \right)}{\left(\sum_{i=1}^{n} E\left((X_i - \bar{x}_i)^2 \right) \right)^{2+\beta}} = 0 \tag{2.59}$$

is satisfied for some $\beta > 0$, then the central limit theorem holds [Adams 1974].

2.2.4.3 Poisson Distribution and Photon Noise

Poisson statistics arise as an inevitable consequence of statistical independence in counting processes. A Poisson distribution also occurs in spite of statistical dependence when the events being counted are rare. Thus, the Poisson distribution is often known as *the distribution of rare events*. Since human perception treats undesirable events as

rare (or else these events would be the norm and not considered to be undesirable as we would have adapted to them), Poisson statistics can often be used to describe undesirable events, for example: occurrence of earthquakes, potholes in roads, defects in factory production lines, mis-dialled phone numbers and dead flies in a soup. The Poisson random variable describes the total number of occurrences of these discrete events in some given continuous interval. The events can occur at any time and the total number of occurrences is not bounded in theory. For a series of events to be modelled by a Poisson distribution, the following conditions must hold.

1. The number of events in any time interval is statistically independent of the number of events in any other non-overlapping interval.

2. The probability of a single event in a small time interval ΔT is proportional to the magnitude of ΔT.

3. The probability of more than one event in a vanishingly small time interval is zero, i.e. events cannot occur simultaneously.

Because it describes a counting process, the Poisson random variable X is discrete. The distribution has only one parameter $\bar{x}$ that describes the mean number of counts. The PDF for a Poisson random variable is given by [Goodman 1985]

$$p_X(x) = \mathrm{Po}\left(\bar{x}\right) = \frac{1}{x!}\bar{x}^{\,x}\exp\left(-\bar{x}\right), \tag{2.60}$$

where x is an integer and the distribution has the property $E(X) = Var(X) = \bar{x}$.

For applications in this thesis, the counting process of interest is photons impinging on a detector. In a *photon-limited* system, the only source of uncertainty is caused by variations in measurements due to the random arrival of photons. Such a source of noise is referred to as *photon noise* and is synonymous with *shot noise* where the former results from counting photons and the latter from counting electrons. Let x_i be the true value of a signal impinging on the ith pixel of a detector. Due to fluctuations in the number of incoming photons, the signal registered on the pixel from one measurement is $\hat{x}_i$ where $\hat{x}_i \sim \mathrm{Po}(x_i)$. The photon noise can thus be expressed as $n_i = \hat{x}_i - x_i$ which has zero mean. Since x_i is strictly positive, the SNR can be defined as

$$\mathrm{SNR} = \frac{\sum_i \hat{x}_i}{\sqrt{\sum_i n_i^2}}, \tag{2.61}$$

where the sum of the squares and square root operation is required as n_i may be negative. Because n_i has zero mean, Eq. (2.61) can be written as $E(\hat{x}_i)/\sigma(n_i)$, in which case

$$\mathrm{SNR} = \frac{E(\hat{x}_i)}{\sigma(\hat{x}_i)}, \tag{2.62}$$

since $\sigma^2(n_i) = \sigma^2(\hat{x}_i - x_i) = \sigma^2(\hat{x}_i)$ by the invariance of the variance under a constant shift. Equation (2.62) says that for a collection of strictly positive measurements, the SNR can be defined as the ratio of their mean and standard deviation. For photon noise, we therefore have

$$\text{SNR} = \frac{x_i}{\sqrt{x_i}} = \sqrt{x_i}. \tag{2.63}$$

Thus, the property that the variance is equal to the expectation for a Poisson distribution implies that photon noise is signal-dependent and the resultant SNR grows as the square root of the signal, or the number of photons captured. Consequently, the only way to reduce photon noise is to capture more signal.

A Poisson distribution with a large mean can be well-approximated by a Gaussian distribution with a variance equal to the mean [Barton 2002], i.e.

$$\lim_{\bar{x}\to\infty} \text{Po}\left(\bar{x}\right) = \mathcal{N}\left(\bar{x}, \bar{x}\right). \tag{2.64}$$

A mean of greater than around 20 usually suffices in practice for the approximation to be used, as values of Poisson probabilities are not usually tabulated for $\bar{x} > 20$ [Barton 2002]. Because a discrete distribution is being approximated by a continuous one, a continuity correction where ± 0.5 is added to the value of an integer Poisson random variable should be applied.

2.3 COHERENT X-RAY DIFFRACTION

The physics of waves and the process of X-ray diffraction are now briefly described. One of the most important properties of a propagating radiation is its coherence.

2.3.1 Coherence

Coherence is required if any meaningful addition of the diffracted waves (resulting in interference fringes) is to be observed in a diffraction experiment. Coherent sources emit radiation with a correlated phase relationship and such relationships allow constructive and destructive interference of wavefronts to occur at specific places. The phase relationship can arise at different times and at different locations, the former being referred to as *temporal* or *longitudinal* coherence and the latter *spatial* or *transverse* coherence. A mixture of both is usually observed in practical waves. In terms of electromagnetic radiation, a wave is temporally coherent when correlations between the phases of the electric field exist at the same location but at different times. A wave is spatially coherent when correlations between the phases of the electric field exist at different locations but at the same time.

The concept of coherence can be characterised by considering two waves each with a different frequency and propagation direction. The *temporal coherence length* of a wave,

L_T, is defined as the distance that the wave must travel before it becomes completely out of phase with a reference wave of some fixed frequency, i.e. when their phase differs by π radians. From Fig. 2.1(a) it can be seen that $2L_T = N\lambda = (N+1)(\lambda - \Delta\lambda)$, where λ is the wavelength of the reference wave, $\Delta\lambda$ is the difference between the wavelength of the two waves and N is the number of wave crests needed for the two waves to become in phase again. If $\Delta\lambda$ is small, then $N \approx \lambda/\Delta\lambda$, giving

$$L_T \approx \frac{\lambda^2}{2\Delta\lambda}. \tag{2.65}$$

The *spatial coherence length* of a wave, L_S, is defined as the distance along a wavefront before that wavefront becomes completely out of phase with a reference wave of a fixed propagation direction. From Fig. 2.1(b) it is evident that $\tan\Delta\theta = \lambda/2L_S$ where $\Delta\theta$ is the discrepancy between the direction of propagation of the two waves, and if $\Delta\theta$ is small,

$$L_S \approx \frac{\lambda}{2\Delta\theta}. \tag{2.66}$$

For waves in more than one dimension, the spatial coherence length becomes the spatial coherence area or volume etc.

A quantum laser with a small spectral spread exhibits high temporal coherence while the high directionality of all laser sources is the result of their high spatial coherence.

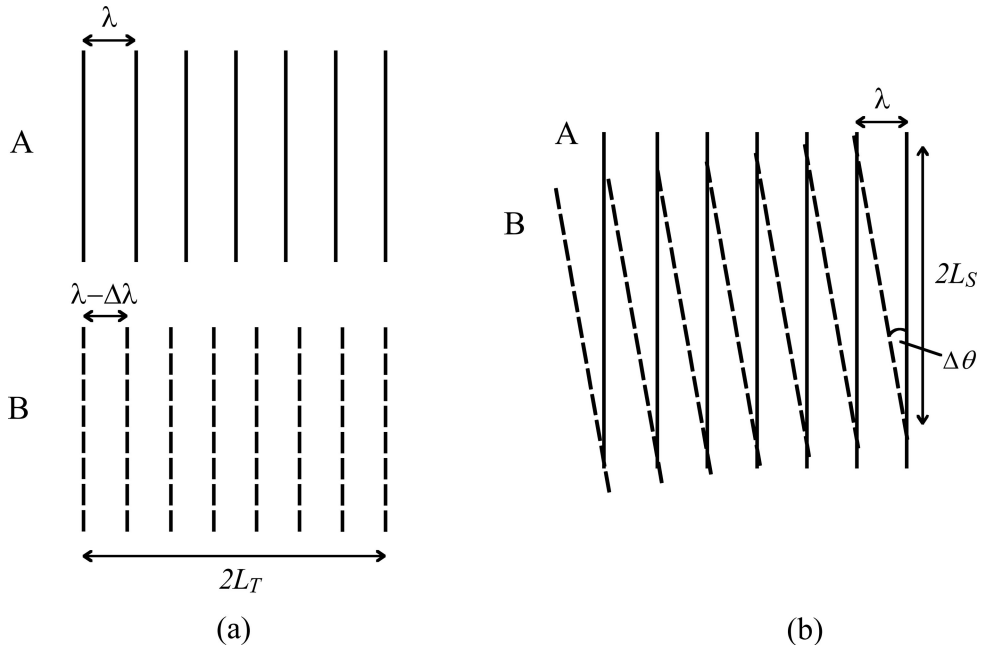

Figure 2.1 Coherence concepts in a propagating wave, adapted from Als-Nielsen and McMorrow [2011]. (a) Temporal coherence. (b) Spatial coherence.

2.3.2 Characteristics of X-ray Sources

Sources of X-rays differ in several respects that include the number of X-ray photons emitted per unit time, the collimation of the emitted beam, the variation in the emitted X-ray wavelength, and the surface area of the source. The collimation of the beam describes how much the beam diverges as it propagates through space and is given in units of milliradians (mrad) both for the horizontal and vertical direction. To characterise the intensity of an X-ray beam, four quantities are commonly used that progressively take into account each of the parameters mentioned above. The four measures are all different normalisations of the number of photons emitted by the source per second (typical units are indicated in brackets):

1. Flux: the total number of X-ray photons passing through a plane crossing the X-ray beam per unit time (photons/s).

2. Fluence: the total number of X-ray photons passing through a unit area of the X-ray beam cross-section per unit time, equivalent to flux density (photons/s/mm^2).

3. Brightness: the total number of X-ray photons passing through a unit solid angle per unit time (photons/s/mrad^2) [2]. Brightness is an appropriate parameter for characterising point sources and for comparing two sources of the same focal spot size.

4. Brilliance: the total number of photons passing through a unit cross-sectional area of the beam within a unit solid angle per unit time (photons/s/mm^2/mrad^2). Brilliance is the most versatile measure to use when comparing the intensity of different X-ray sources.

Normalisation against variations in the emitted wavelength, or the spectral distribution of the X-rays, is also important because some sources may produce very smooth spectra while others may contain peaks at certain photon energies, so it is sensible to specify the range of the photon energy when quoting the above measures. The convention is to define the photon energy range normalisation as 0.1% of the total bandwidth [He 2009, Als-Nielsen and McMorrow 2011].

2.3.3 Diffraction by a General Object

Consider a macroscopic object through which there sweeps a train of monochromatic planar X-ray waves. Excited by the impinging X-rays, each electron in this object becomes the centre of a spherical scattering wavefront. It is then of interest to find the combined effect of the interference from all these scattered wavefronts at a point

[2] The SI unit for solid angle is the steradian (sr) but mrad^2 is commonly used throughout the X-ray light source literature.

outside this object. The problem of diffraction in its most general form is difficult (see Born and Wolf [2003] for an in-depth treatment); here, the problem is reduced by making several simplifying assumptions.

First, assume that the detector is placed a large distance away from the object, or equivalently, that the object's linear dimension is small with respect to the detection distance. This is the *Fraunhofer* or *far-field approximation* which amounts to saying that two paths to the same point in the detection plane from two different regions in the scattering object are parallel. Second, assume that multiple scattering where the scattered energy is rescattered does not take place. Furthermore, absorption for both the incident and scattered radiation is considered to be minimal and ignored. These assumptions form the *Born approximation*. Finally, assume the speed of the X-ray wavefront that travels through the object is always constant and that interactions between incident and scattered waves are neglected. These numerous unrealistic-sounding assumptions might not seem very sensible, but in practice fairly accurate results can nevertheless be obtained while allowing the analysis to be greatly simplified.

Let the object be described by a scattering density function, $f(\mathbf{x})$. In the case of X-rays it is the electrons inside the material that interact with the incident radiation and $f(\mathbf{x})$ is therefore the electron density of the object. Now consider two volume elements A_1 and A_2 in the diffracting object as shown in Fig. 2.2. The incoming planar X-rays have a direction given by the unit vector $\mathbf{s}_i$ and we consider the outgoing X-rays after scattering in the direction given by the unit vector $\mathbf{s}_o$. If we choose, without loss of generality, a coordinate system whose origin is centred at the volume element A_1, then we can construct a vector $\mathbf{x}$ that gives us the relative position of A_2 with respect to A_1. The combined effect of the scattered X-rays from elements A_1 and A_2 is governed by their path difference

$$\overline{A_1C} - \overline{A_2B} = \mathbf{x} \cdot (\mathbf{s}_o - \mathbf{s}_i) \,. \tag{2.67}$$

The phase difference ϕ that results is then the ratio of the path difference to the wavelength λ of the X-ray normalised to 2π, giving

$$\phi \left(\mathbf{x}, \mathbf{s}_i, \mathbf{s}_o\right) = k\mathbf{x} \cdot (\mathbf{s}_o - \mathbf{s}_i) \,, \tag{2.68}$$

where $k = 2\pi/\lambda$ is the wave number.

The strength of the scattering from a three-dimensional element, forming the amplitude of the scattered wavefront, is given by the product of the scattering density with the volume of that element, i.e. $f(\mathbf{x})\Delta\mathbf{x}$, where $\Delta\mathbf{x}$ is the elemental volume. Hence, the time-independent electric field resulting from the interference of these two scattered X-ray wavefronts can be written as

$$F_{A_1A_2} = f\left(\mathbf{0}\right)\Delta\mathbf{x} + f\left(\mathbf{x}\right)\Delta\mathbf{x}\exp\left(-i\phi\left(\mathbf{x}, \mathbf{s}_i, \mathbf{s}_o\right)\right) . \tag{2.69}$$

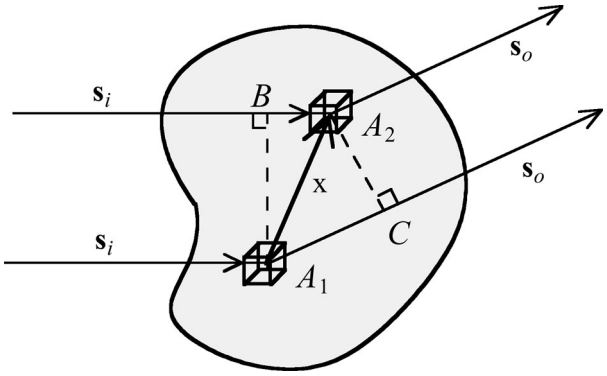

Figure 2.2 Diffraction of two volume elements from an arbitrary three-dimensional object.

Summing over the contribution from all volume elements inside the object gives the overall electric field of the diffracted X-ray seen from a point on the far-field detector plane as a function of the phase difference ϕ

$$F(\phi) = \sum_{\mathbf{x}} f(\mathbf{x}) \Delta\mathbf{x} \exp(-i\phi(\mathbf{x}, \mathbf{s}_i, \mathbf{s}_o)), \tag{2.70}$$

which via Eq. (2.68) is equal to

$$F(\phi) = \sum_{\mathbf{x}} f(\mathbf{x}) \exp(-ik\mathbf{x} \cdot (\mathbf{s}_o - \mathbf{s}_i)) \Delta\mathbf{x}. \tag{2.71}$$

Letting the volume elements shrink to infinitesimal size, the sum becomes an integral and we have

$$F(\mathbf{u}) = \int_{-\infty}^{\infty} f(\mathbf{x}) \exp(-i2\pi\mathbf{u} \cdot \mathbf{x}) \, \mathrm{d}\mathbf{x}, \tag{2.72}$$

where the scattering vector $\mathbf{u}$ is defined by

$$\mathbf{u} = \frac{\Delta\mathbf{s}}{\lambda}, \tag{2.73}$$

and $\Delta\mathbf{s} = \mathbf{s}_o - \mathbf{s}_i$. Equation (2.72) is exactly the form of a Fourier integral, thus the far-field diffraction pattern can be thought of as the Fourier transform of the scattering density. The physical quantity usually measured in a diffraction experiment is the intensity of the diffracted field, defined as $|F(\mathbf{u})|^2 = F(\mathbf{u}) F^*(\mathbf{u})$.

2.3.4 Diffraction by a Crystal

Diffraction by a single crystal is presented here for completeness and to establish terminologies used in crystallography. It will be treated in more detail in Chapter 4.

A crystal of any material is a three-dimensional periodic arrangement of the molecules of that material. When the material crystallises, its molecules attempt to reach the lowest free-energy state. This is often accomplished by the molecules packing them-

selves in a regular way. It is interesting that even large protein molecules with millions of atoms are able to follow this rule [Drenth 2007].

Consider now the diffraction of a crystalline object. Since a crystal is a spatially periodic, translational repetition of a single aperiodic motif in 3-D, it can be described by defining three vectors $\mathbf{a}_1$, $\mathbf{a}_2$, $\mathbf{a}_3$, called the *cell constants*, and the aperiodic motif is referred to as the *unit cell*. The most general three-dimensional crystal geometry is that of a parallelepiped unit cell. Using the cell constants, any unit cell can be located via three unique integers m_1, m_2 and m_3 giving a triplet $\mathbf{m} = [m_1, m_2, m_3]^T$ that allows the position $\mathbf{x_m}$ of any unit cell within the crystal to be specified by

$$\mathbf{x_m} = m_1\mathbf{a}_1 + m_2\mathbf{a}_2 + m_3\mathbf{a}_3 = \mathrm{L}\mathbf{m}, \tag{2.74}$$

where as before in Sec. 2.1.4, L is the generator matrix of a lattice and it can be written in this case as $\mathrm{L} = [\mathbf{a}_1|\mathbf{a}_2|\mathbf{a}_3]$ such that its columns are the cell constants. For a finite crystal with a total of $N_1 \times N_2 \times N_3$ unit cells, where N_1, N_2 and N_3 are the number of unit cells in the direction of each of the cell constants, respectively, m_i takes on only values $0, \cdots, N_i - 1$ for all $i = 1, 2, 3$. The electron density of the crystal, denoted by $g(\mathbf{x})$, is then given by

$$g(\mathbf{x}) = \sum_{\mathbf{m}} f(\mathbf{x} - \mathbf{x_m}) = f(\mathbf{x}) \otimes s(\mathbf{x}), \tag{2.75}$$

where

$$s(\mathbf{x}) = \mathrm{comb}(\mathrm{L}^{-1}\mathbf{x}) \tag{2.76}$$

is called the *shape function* of the crystal.

The diffracted field by a crystal, $G(\mathbf{u})$, is the coherent sum of the diffraction by each unit cell. By Eq. (2.71), each unit cell introduces an additional phase shift $2\pi\mathbf{u}\cdot\mathbf{x_m}$, and since all unit cells are identical,

$$G(\mathbf{u}) = F(\mathbf{u}) \sum_{\mathbf{x_m}} \exp\left(-i2\pi\mathbf{u}\cdot\mathbf{x_m}\right), \tag{2.77}$$

where $F(\mathbf{u})$ denotes the scattering from one unit cell. Expanding $\mathbf{x_m}$ out along the components of the crystal lattice via Eq. (2.74), the diffracted field from a crystal becomes

$$G(\mathbf{u}) = F(\mathbf{u}) \prod_{j=1}^{3} \left(\sum_{m_j=0}^{N_j-1} \exp\left(-i2\pi m_j \mathbf{a}_j \cdot \mathbf{u}\right) \right). \tag{2.78}$$

Each sum is a geometric series and it can be readily shown that

$$G(\mathbf{u}) = F(\mathbf{u}) \prod_{j=1}^{3} \left(\frac{\sin\left(N_j\Psi_j\right)}{\sin\Psi_j} \exp\left(-i\left(N_j - 1\right)\Psi_j\right) \right), \tag{2.79}$$

where

$$\Psi_j = \pi \mathbf{a}_j \cdot \mathbf{u}. \tag{2.80}$$

Letting the summation variable m_j in Eq. (2.78) extend from $-\infty$ to $+\infty$ models the situation for an infinite crystal, which is the default assumption in traditional X-ray crystallography.

The intensity of the diffraction by a crystal is then

$$|G(\mathbf{u})|^2 = |F(\mathbf{u})|^2 \prod_{j=1}^{3} \left(\frac{\sin^2 (N_j \Psi_j)}{\sin^2 \Psi_j} \right). \tag{2.81}$$

In Chapter 4, Eq. (2.81) will be written more compactly as

$$I(\mathbf{u}) = |F(\mathbf{u})|^2 S^2(\mathbf{u}), \tag{2.82}$$

where $I(\mathbf{u}) = |G(\mathbf{u})|^2$ and $S(\mathbf{u}) = \prod_{i=1}^{3} \sin(N_i \Psi_i) / \sin(\Psi_i)$ is called the *transform of the shape function* as it is the Fourier transform of the finite lattice that specifies the location of each unit cell that in turn gives the crystal its shape.

2.3.5 The Reciprocal Lattice and the Ewald Sphere

Fourier space, referred to as the *reciprocal space* in crystallography, is also spanned by three basis vectors: $\mathbf{b}_1$, $\mathbf{b}_2$ and $\mathbf{b}_3$. These vectors are related to the cell constants by the equations [Drenth 2007]

$$\mathbf{b}_1 = \frac{\mathbf{a}_2 \times \mathbf{a}_3}{\mathbf{a}_1 \cdot (\mathbf{a}_2 \times \mathbf{a}_3)} \tag{2.83}$$

$$\mathbf{b}_2 = \frac{\mathbf{a}_3 \times \mathbf{a}_1}{\mathbf{a}_2 \cdot (\mathbf{a}_3 \times \mathbf{a}_1)} \tag{2.84}$$

$$\mathbf{b}_3 = \frac{\mathbf{a}_1 \times \mathbf{a}_2}{\mathbf{a}_3 \cdot (\mathbf{a}_1 \times \mathbf{a}_2)}, \tag{2.85}$$

and together define a counterpart to the real lattice of the crystal called the *reciprocal lattice* in reciprocal space. The reciprocal lattice points in reciprocal space is denoted by the vector $\mathbf{u_h}$ and can be written as

$$\mathbf{u_h} = h_1 \mathbf{b}_1 + h_2 \mathbf{b}_2 + h_3 \mathbf{b}_3 = \mathrm{L}^{-1} \mathbf{h}, \tag{2.86}$$

where h_1, h_2, h_3 are integers and referred to as the *Miller indices*[3], $\mathbf{h} = [h_1, h_2, h_3]^T$ and $\mathrm{L}^{-1} = [\mathbf{b}_1 | \mathbf{b}_2 | \mathbf{b}_3]$ is the inverse of the lattice matrix for the real space crystal lattice. The unit cell of a crystal lattice in real space and the unit cell of its corresponding lattice in reciprocal space is shown in Fig. 2.3.

[3]The Miller indices are usually denoted by h, k, l in the literature.

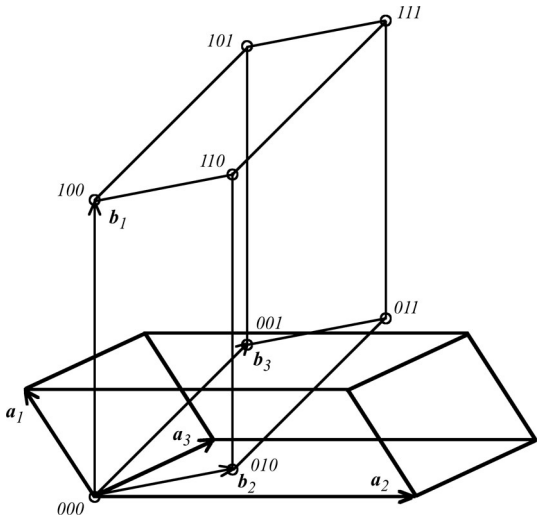

Figure 2.3 A real (thick line) and the corresponding reciprocal (thin line) lattice unit cell. Miller indices are labelled on the edges of the reciprocal lattice.

Referring to Eq. (2.79), a crystal of infinite extent will only give non-zero diffraction at reciprocal space locations where Ψ_j is an integer multiple of π, i.e., diffraction of an infinite crystal will only occur at the reciprocal lattice points $\mathbf{u_h}$. This statement is referred to as the *Laue conditions* which must be satisfied for the diffraction by an infinite crystal to be observable. Physically, this can be explained by noting that the diffracted beams of X-rays will only emerge from a crystal when constructive interference occurs between the X-rays reflected from all successive parallel inter-crystal planes. For an incoming X-ray beam of incidence angle θ on a series of crystal planes separated by spacing $d_{\mathbf{h}}$, it can be shown that the condition for constructive interference of all outgoing wavefronts is given by

$$m\lambda = 2d_{\mathbf{h}} \sin\theta \tag{2.87}$$

for m any integer, known as *Bragg's Law*. The Laue condition can be shown to be the vector equivalent of Bragg's law [James 1948] and the peaks resulting from the diffraction maxima of an infinite crystal are akin to delta functions and are referred to as *Bragg peaks* or *Bragg reflections*.

In the regime of elastic scattering, where only the X-ray photons that maintain their energy after scattering are of interest, the Fourier space position $\mathbf{u}$ can be thought of as the momentum transfer vector between the incoming and outgoing X-ray photons. The condition for constructive interference can thus be demonstrated in reciprocal space by applying the law of conservation of momentum, resulting in a sphere traced out by the outgoing scattering vector with its surface passing through the origin of reciprocal space as shown in Fig. 2.4. This geometrical description of diffraction was proposed by Ewald [Ewald 1969] and encompasses both Bragg's law and the Laue conditions.

The Ewald construction, known as the *Ewald sphere*, allows the determination of the Bragg reflections that will be observed if the crystal orientation, and hence the reciprocal lattice orientation, with respect to the incident X-ray beam is known. Physically rotating the crystal amounts to rotating the reciprocal lattice about its origin. For all instances where the Ewald sphere intersects a reciprocal lattice point, Bragg's law is obeyed and a diffraction maximum will occur. Drawing the scattering vectors on a diagram such as that shown in Fig. 2.4, it can be seen that the magnitude of the reciprocal space position vector $\mathbf{u} = \Delta\mathbf{s}/\lambda$ must be between zero and $2/\lambda$. Rotating the crystal through all possible orientations therefore allows the outermost point on the Ewald sphere to trace out another spherical surface with radius $2/\lambda$ known as the *resolution sphere*. The geometric construction of the resolution sphere visualises the limit of all observable Bragg reflections in reciprocal space for a given X-ray wavelength λ, indicating the highest obtainable resolution from the experiment.

In practice, since real crystals have a finite size and are slightly imperfect, the Bragg reflections have a finite size also, and it is sufficient that the reciprocal lattice point is close to the Ewald sphere for diffraction to be observed.

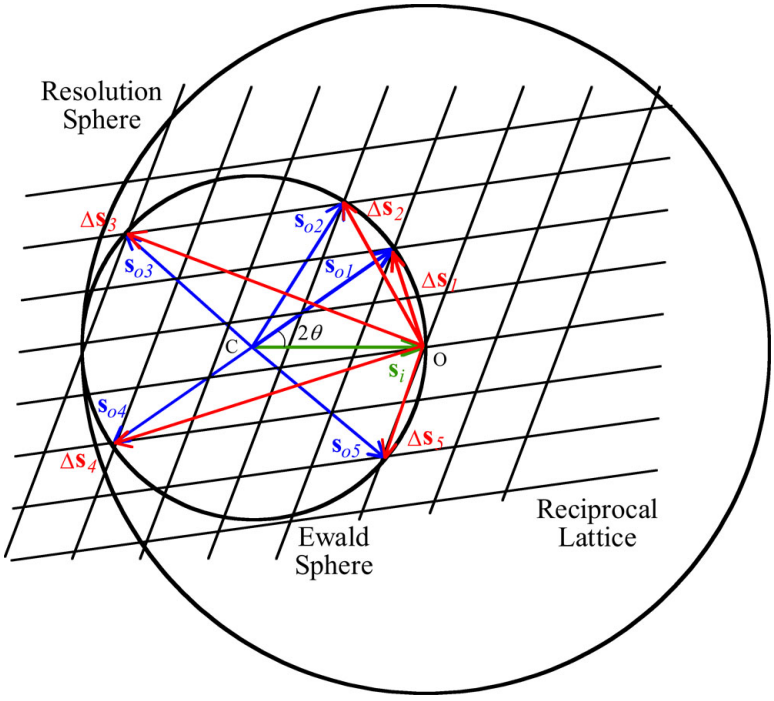

Figure 2.4 A 2-D slice of the Ewald sphere construct in 3-D reciprocal space. The incoming (green) and outgoing (blue) scattering vectors are of the same length because of the assumption that photon momentum is conserved. Five Bragg reflections corresponding to the difference between the incoming and outgoing scattering vectors (red) can be observed in this plane. The crystal is situated at C. Rotating the crystal corresponds to rotation of the reciprocal lattice about its origin O.

2.4 SPACE GROUPS AND SYMMETRY

The space groups of crystal structures are akin to the major and minor keys in musical composition using the equal tempered scale. It is a way of classifying a collection of objects according to a set of rules. There are 12 major and 12 minor keys in music theory and there are 230 space groups in crystallography in 3-D. The set of movements required to restore an object to its original state is its symmetry. The classification of these movements in space is described by the space groups.

Two-dimensional space groups are often encountered in everyday life from elaborate mosaics and frieze artworks to bathroom tiles, carpet designs and simple brick walls. See the beautifully illustrated book by Wade [2006] for many more examples. Space groups in crystallography enumerate the collection of symmetry operators that describe an infinite crystal. The definitive reference is the International Tables for Crystallography [Hahn 2005].

2.4.1 The Asymmetric Unit

In practice, most crystal unit cells will be composed of several identical molecules arranged in some fashion that produces symmetry. The unit cell is a collection of molecules that can be translated (at least once) to build up the crystal. The largest collection of molecules in a crystal that possess no symmetry elements itself but can be *both* rotated and translated to build up the crystal is called the *asymmetric unit*. Note the emphasis that only translation is allowed for the unit cell to build up a crystal, but both translation and rotations are allowed for the asymmetric unit to do the same.

The asymmetric unit embodies the minimum amount of information that is needed to describe the entire crystal. In the simplest case, the asymmetric unit is a single molecule. Much more will be said in Chapter 6 on the possibility of having multiple different unit cells to describe the same crystal when crystals become small.

2.4.2 Groups

A *group* is a set G together with a binary operation $\circ$ that combines two elements, x and y, to give a third element $x \circ y$, such that the four conditions below are satisfied

1. Closure: $x \circ y \in G$ for all $x, y \in G$.
2. Associativity: $(x \circ y) \circ z = x \circ (y \circ z)$ for all $x, y, z \in G$.
3. Identity: There exists an element $e \in G$ such that $e \circ x = x \circ e = x$ for all $x \in G$. The element e is called the identity.
4. Inverse: There exists an element $y \in G$ such that $x \circ y = y \circ x = e$ for all $x \in G$. The element y is called the inverse of x.

2.4.3 Space Groups

Space groups are collections of symmetry operations that form a group and is consistent with translational symmetry. A symmetry operation of a given object in a given space is a motion in that space that maps the object onto itself. The motions allowed in 3-D are translations, rotations, reflections about a point, reflections about a plane, and combinations of any of the above. For example, a translation plus a rotation gives rise to motions called *screw axes*. A reflection plus a shift is called a *glide*. A rotation about an axis plus a reflection in a plane gives rise to an *improper rotation*, also known as *rotoreflection*.

The general formula for the action of an element of a space group that transforms a point $\mathbf{x}$ into a new point $\mathbf{y}$ is given by

$$\mathbf{y} = \mathrm{M}\mathbf{x} + \mathbf{d}, \tag{2.88}$$

where M is a matrix describing the motion, $\mathbf{d}$ is a column vector that is a function of the motion and the particular crystal lattice the point $\mathbf{x}$ is situated in (14 lattices are possible and are known as the Bravais lattices). The matrices M form a point group that is a basis of the space group.

The *International Symbol*, also known as the *Hermann-Mauguin* notation, is commonly used to notate the space groups. A capital letter P, I, F, A, B, C, or R is used to denote the lattice centreing, and the numerical symbols that follow denote the symmetry operations that can be done on the unit cell without changing its appearance. The numerical symbols describe the most prominent symmetry operation visible when projected along one of the symmetry directions of the crystal.

For example, the space group for crystals of the so-called Photosystem I protein in its trimeric form is $\mathrm{P6_3}$ [Jordan et al. 2001], meaning that each unit cell of the crystal exhibits primitive centring (i.e., one net asymmetric unit per unit cell), with the symbol 6_3 denoting a translation of half a unit cell along one unit cell axis and a 180° rotation about the plane formed by the other two unit cell axes (rotating three times about a six-fold screw axis).

Space groups in other dimensions are also possible. There are 2 space group in 1-D (the trivial space group P1 and one with a reflection point $\mathrm{P\bar{1}}$), 17 space groups in 2-D (called the plane symmetry group or the wallpaper group) and 230 space groups in 3-D. Groups of symmetries whose translational behaviour is restricted to a dimension lower than that of the space on which the symmetry operations act are known as *subperiodic groups*. Some examples of subperiodic groups include frieze groups (groups in a plane with one-dimensional translations), rod groups (groups in 3-D space with one-dimensional translations) and layer groups (groups in 3-D space with 2-D translations).

The symmetry elements found in protein crystals are those composed of only trans-

lations and rotations. This is because biological molecules are composed of chiral amino-acid residues in nature, which are further coiled into larger structures that are also chiral - such as the right-handed helices of DNA [Lavelle 2009] [4]. In total, the disallowance for mirror elements reduces the number of space groups for biological molecules from 230 to 65.

Figure 2.5 shows all space groups in 2-D consisting purely of rotational symmetry elements, applicable in describing 2-D crystals (i.e., one layer of 3-D molecules) often encountered in electron crystallography. There are 5 such groups altogether in 2-D.

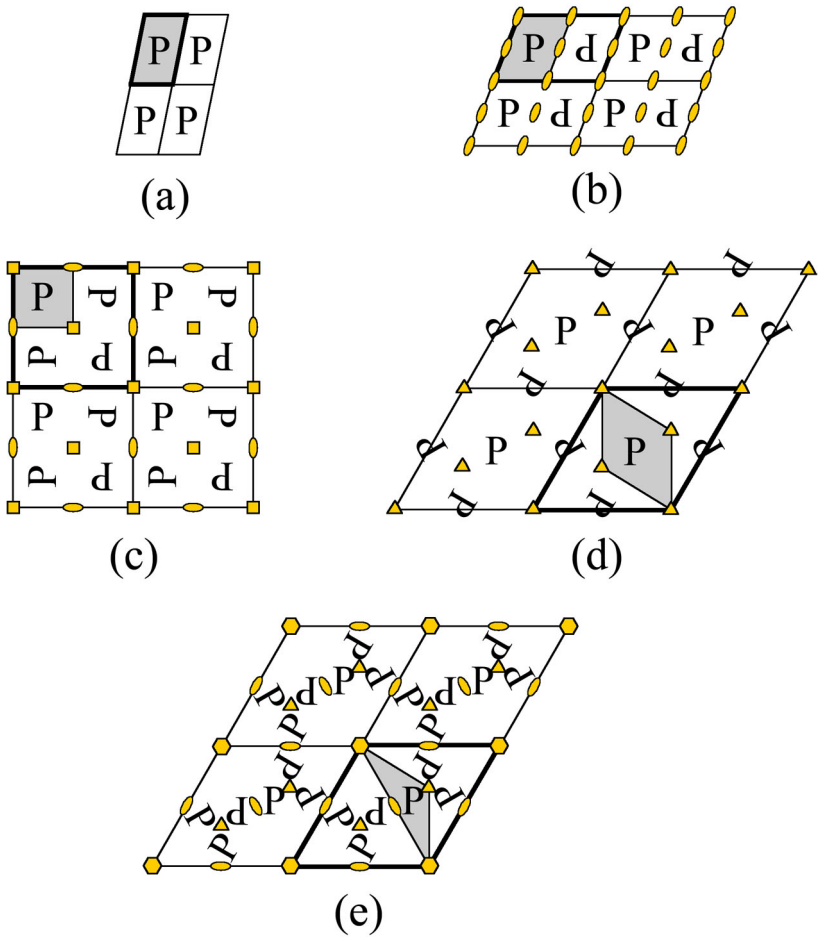

Figure 2.5 Two-dimensional space groups consisting of only rotational elements. (a) P1, (b) P2, (c) P4, (d) P3 and (e) P6. The asymmetric units are shaded in grey and the unit cells are emphasised by a thick bounding line.

[4]DNA molecules are subdivided into three classes, A-DNA, B-DNA and Z-DNA. A- and B-DNAs are right-handed while Z-DNA is in fact left-handed.

2.5 ITERATIVE PROJECTION ALGORITHMS

Many problems can be stated in the form of a constraint satisfaction problem. Formulated in this way, any knowledge about the problem can be viewed as a constraint such that the known information are subsets of a multidimensional space consisting of all possibilities. The locations where all of the information are consistent, i.e. the intersection of those sets, are the solutions to the initial problem. The idea behind iterative projection algorithms (IPAs) is to start at a point in this multidimensional space, and attempt to traverse it to arrive at the constraint set intersection.

It is convenient to formulate IPAs as operations on vectors in an n-dimensional metric space. A vector $\mathbf{x} = [x_1, \cdots, x_n]^T$ in this space can represent an n-pixel image where each of its n components corresponds to the value of each of the n pixels of the image. Different choices of metric give rise to different metric spaces. The multidimensional aspect of these problems where each variable can be seen as a separate dimension makes the search difficult but can also give rise to structures that provide additional information about the nature of the problem.

A solution may not exist (constraints may not intersect) or multiple solutions may exist depending on the nature of the constraint. IPAs in general accept only two constraints by default. For problems where multiple constraints need to be satisfied, the constraints would either have to be combined, or if that is not possible, an approach such as the "divide and concur" method [Gravel and Elser 2008] where constraints are individually satisfied and the result brought together by averaging, can be used to allow IPAs to be applied to problems with more than two constraints.

2.5.1 Distance, Constraints and The Landscape

The dynamics of a given iterative projection algorithm depend entirely on the constraints. Constraint sets can be characterised by their convexity. A set is *convex* if all points on a line between any two points in the set are also in the set. More precisely, if $\mathbf{x}$ and $\mathbf{y}$ are any elements of a convex constraint set C then

$$\mathbf{x} + \eta(\mathbf{y} - \mathbf{x}) \in C \tag{2.89}$$

for all $0 \leq \eta \leq 1$. A *non-convex* set has points in it that do not satisfy the above condition.

The convexity of a constraint is directly related to the difficulty of the constraint satisfaction problem. Satisfying one constraint of any convexity is easy. Satisfying more than one constraint where each of the constraints is convex is also straightforward. But as soon as one of the constraints is non-convex in a problem with many constraints to be satisfied, the problem becomes non-trivial. This statement can be illustrated by first considering the distance between a point and a constraint set. Write the distance

between $\mathbf{x}$ and $\mathbf{y}$ as $||\mathbf{y} - \mathbf{x}||$. The usual Euclidean norm is employed in this case such that

$$||\mathbf{y} - \mathbf{x}|| = \sqrt{\sum_{j=1}^{n} (y_j - x_j)^2}. \tag{2.90}$$

Calculating the sum of the minimum distance from a point in the search space to all individual constraint sets in that space gives a value which can be interpreted as the "height" above ground in the "landscape" that the IPA will have to traverse to find the solution[5]. This so-called landscape is referred to herein more formally as the *minimum-distance landscape* and a more precise mathematical definition is given in Eq. (2.92) in the next section after the notion of projections has been developed.

The hills in the minimum distance landscape are locations far away from at least one of the constraint sets and the valleys are locations close to all constraints. The position(s) where the height is zero is(are) where the solution(s) lie. Valleys of the landscape that do not reach zero height, form what are known as *false fixed-points* and may cause the traversal of the landscape by an IPA to become trapped. The process of trapping is referred to as *stagnation*.

It makes sense to think of a false fixed-point as a local minimum so the difficulty of a constraint satisfaction problem can be characterised by the number of local minima there are. Two convex constraints are illustrated using rectangles in Fig. 2.6(a) and a non-convex constraint is shown in Fig. 2.6(b). In Fig. 2.6(c), two highly non-convex constraints are shown which consist of uniformly randomly distributed points that intersect at two disjoint locations (see if you can spot them). The minimum-distance landscape of these three scenarios is shown on the bottom row of the same figure. The extreme difference in the number of local minima between the convex problem and the structureless, random, non-convex one is visibly apparent by their corresponding search landscapes. It is clear from these figures that convexity ensures that there is only a single unique minimum-distance region in the search space.

The most straightforward method to seek out the solution of a constraint satisfaction problem is to enumerate the entire search space and carry out a brute-force search to systematically interrogate each point in the space. However this method is impractical even for small problems. An efficient and simple approach is to move from one point in the search space to another in such a way that individual moves are based on information garnered from the previous location. IPAs are designed to tackle constraint satisfaction problems in this way. They seek out the intersection between two constraint sets by iteratively searching through the multidimensional space using operators called *projections*. Note that gradient methods can also be applied to traverse the space and has been compared with IPAs [Fienup 1982].

[5]See Susskind [2005] where he advocates the "cosmic landscape" that defines all possible universes that can arise from string theories.

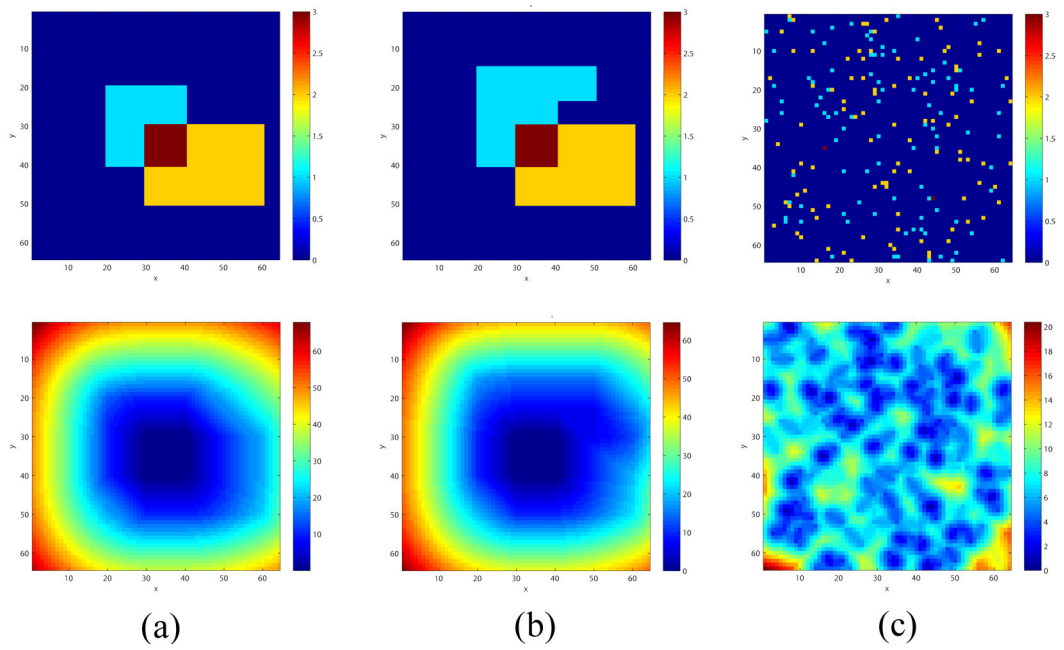

Figure 2.6 Search space (top row) and the corresponding minimum-distance landscape (bottom row) for (a) convex constraints, (b) non-convex constraint, (c) non-convex and random constraints.

2.5.2 Projections

A projection P_A is defined as an operation that takes a point $\mathbf{x}$ in a metric space to the closest point $\mathbf{y} = P_A\mathbf{x}$ on a set A in that space. In notations,

$$\mathbf{y} = P_A\mathbf{x} = \text{argmin}_{\mathbf{y}\in A}||\mathbf{y} - \mathbf{x}||, \tag{2.91}$$

where $\text{argmin}_{\mathbf{x}} f(\mathbf{x})$ denotes the value of $\mathbf{x}$ that minimises $f(\mathbf{x})$. A projection is therefore an operation that makes the minimum change to a signal such that the modified signal satisfies the constraint being projected onto.

Using the projection operator, the concept of the minimum-distance landscape introduced in the previous section can be defined for N constraints, $A_1, \cdots, A_N$, as

$$L(\mathbf{x}) = \frac{1}{N}\sum_{n=1}^{N}||\mathbf{x} - P_{A_n}\mathbf{x}||, \tag{2.92}$$

and the solution to the constraint satisfaction problem is when $L(\mathbf{x}) = 0$.

A generalisation of the projection operation is that of a *relaxed projection*. The stringency that the iterate must move exactly onto the constraint surface is relaxed and leeway is given such that the iterate can land on a location further into or away from the constraint surface. The new location is given by a linear combination of the current position and the projected position. Denoting a relaxed projection onto a constraint

set A by F_A, we can then write

$$F_A\mathbf{x} = P_A\mathbf{x} + \gamma_A(P_A - I)\mathbf{x}, \tag{2.93}$$

where I is the identity operator defined by $I\mathbf{x} = \mathbf{x}$. The amount of relaxation is controlled by the constant γ_A which usually satisfies $|\gamma_A| \leq 1$. A special case of a relaxed projection is the *reflection*,

$$R_A\mathbf{x} = (2P_A - I)\mathbf{x}, \tag{2.94}$$

occurring when $\gamma_A = 1$, which moves a point the same distance into the constraint set as it was outside the set.

2.5.3 Projection Algorithms

Combinations of projection operators form sets of rules that can be used to traverse the search landscape. The algorithms are recursive in nature and different projection algorithms are distinguished by different update strategies for forming a new iterate from the old. At the ith iteration of an IPA, the current iterate $\mathbf{x}_i$ is updated to form the next iterate $\mathbf{x}_{i+1}$ following a function f, written in general as

$$\mathbf{x}_{i+1} = f(\mathbf{x}_i). \tag{2.95}$$

The update rule f of the algorithm is a combination of the projections P_A, P_B and the identity operator I.

Convergence is reached when the iterate is no longer changing, i.e, at a fixed point, defined by $\mathbf{x}_{i+1} = \mathbf{x}_i$, or in general, the iterate cycles around some sort of attractor. If the iterate is not near a solution upon convergence, stagnation has occurred.

2.5.3.1 Error-reduction Algorithm

von Neumann [1950] was among the first to propose the idea of a projection-based algorithm. He realised that the most basic strategy is simply to move back-and-forth between the surface of one constraint set and the other. This is now known as the error-reduction (ER) algorithm in the phase retrieval community and is the simplest conceptual form of IPA. The update rule for the algorithm is

$$\mathbf{x}_{i+1} = P_B P_A \mathbf{x}_i. \tag{2.96}$$

The ER algorithm derives its name from the property that however slowly, the distance from the current iterate to the solution, i.e., the error, always decreases monotonically with increasing numbers of iteration - provided the constraint geometry is convex

[Fienup 1982]. Applied to non-convex constraints, the ER algorithm almost always result in stagnation, as can be seen in Fig. 2.7(b) and (c). The only situation where it is effective is when the iterate is close to a solution.

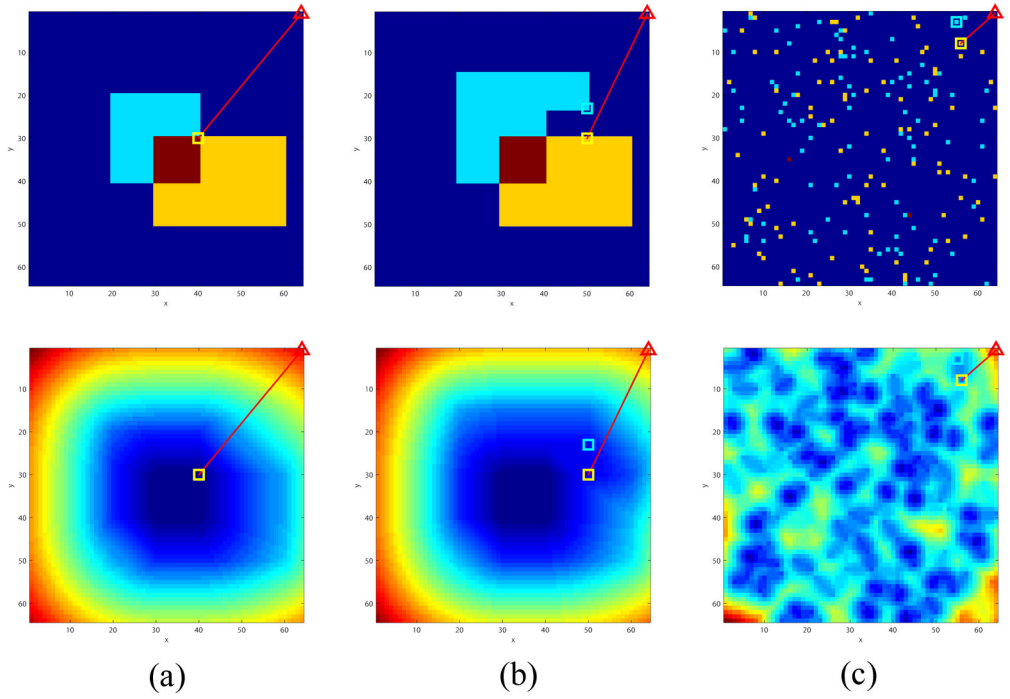

Figure 2.7 Behaviour of the ER algorithm for (a) convex constraints, (b) non-convex constraint, (c) non-convex random constraints. Red path indicates the trajectory of the algorithm.

2.5.3.2 Fienup's Input-Output Algorithm

The price to be paid for the simplicity and monotonic reduction in the error using the ER algorithm is the tendency for the iterate to be trapped at fixed points which are not a solution. In his well-cited 1982 paper, Fienup extended the ER algorithm and proposed the now widely adopted hybrid-input-output (HIO) algorithm based on ideas from nonlinear feedback control theory [Fienup 1982]. The HIO algorithm is given by

$$\mathbf{x}_{i+1} = \begin{cases} P_B \mathbf{x}_i & x_j \in A \\ (I - \beta P_B)\mathbf{x}_i & x_j \notin A. \end{cases} \tag{2.97}$$

for all $j = 1, \cdots, n$. The conditional statement can be rewritten as [Millane and Lo 2013]

$$\mathbf{x}_{i+1} = \left(P_A P_B + P_{A'}\left(I - \beta P_B\right)\right)\mathbf{x}_i, \tag{2.98}$$

where $P_{A'}$ is the projection onto A', the complementary set of A, representing the set of all x_j outside the support.

The HIO algorithm was originally conceived within the mindset of applications in phase retrieval where constraints in real space and Fourier space need to be satisfied. Constraints in real space in those applications are usually knowledge of the support and that the density of the object is positive. Thus it turns out that the HIO algorithm in the form of Eq. (2.97) is suitable only for real space constraints such as finite support or positivity. Millane and Stroud [1997] have generalised the HIO algorithm such that it can accept any constraints.

The performance of HIO for constraints of different convexity is shown in Fig. 2.8, and it can be seen that it is able to escape some local minima and avoids sometimes being trapped.

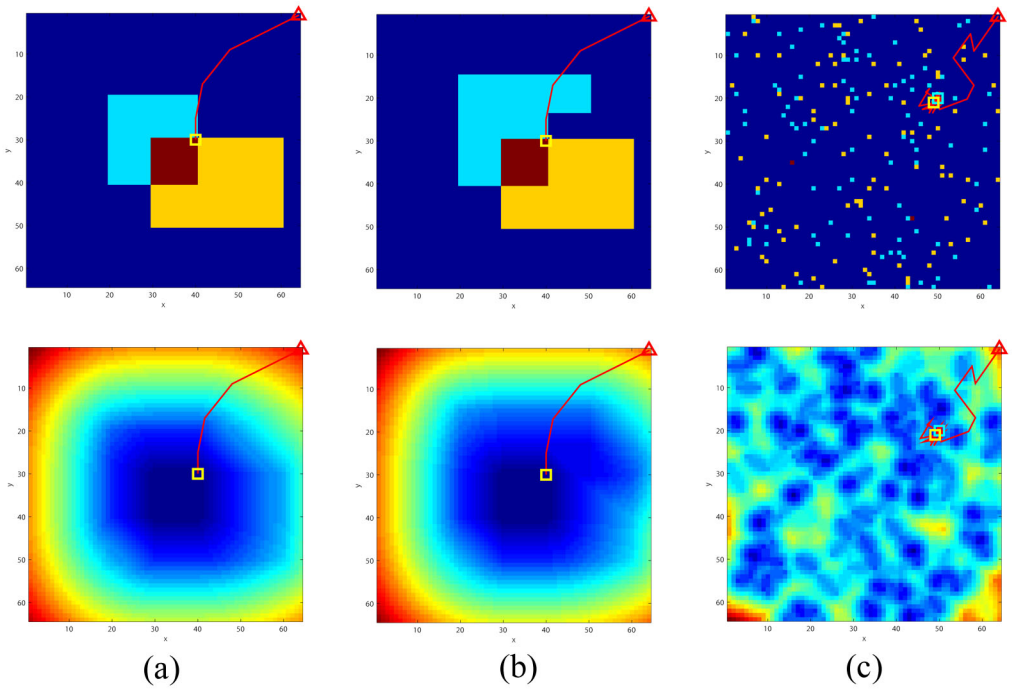

Figure 2.8 Behaviour of the HIO algorithm for (a) convex constraints, (b) non-convex constraint, (c) non-convex random constraints. Red path indicates the trajectory of the algorithm.

2.5.3.3 Elser's Difference Map Algorithm

A more general IPA has been described by Elser [2003b]. This so-called difference map (DM) algorithm involves the use of three independent parameters γ_S, γ_M and β and takes the form

$$\mathbf{x}_{i+1} = \big(I + \beta\,(P_A F_B - P_B F_A)\,\big)\mathbf{x}_i, \tag{2.99}$$

where F_A and F_B are the relaxed projections given in the previous section for two different values of the relaxation parameter γ_A and γ_B. A general rule that works

well in most cases consists of $\gamma_A = -1/\beta$ and $\gamma_B = 1/\beta$ [Elser 2003a]. A property of the DM algorithm is that its fixed points can always be used to directly generate a solution. This can be shown as follows. Suppose a fixed point is found, so $\mathbf{x}_{i+1} = \mathbf{x}_i$, then Eq. (2.99) implies that

$$P_A F_B \mathbf{x}_i = P_B F_A \mathbf{x}_i. \tag{2.100}$$

By noting that each side of Eq. (2.100) satisfies both of the constraints, the solution, $\mathbf{x}_{\text{soln}}$, can then be written as

$$\mathbf{x}_{\text{soln}} = P_A F_B \mathbf{x}_i = P_B F_A \mathbf{x}_i. \tag{2.101}$$

The DM algorithm has good global convergence properties and upon selecting particular values for the independent parameters, this algorithm can take the form of many of the previously mentioned IPAs. The DM algorithm can thus be considered as a generalisation of a large class of IPAs. It is also from the advent of the DM algorithm that IPAs began to emerge from the confines of traditional phasing problems into the wider field of general constraint satisfaction problems [Elser et al. 2007].

Trajectories of a DM algorithm run for constraints of different convexity is shown in Fig. 2.9, and it can be seen that it is able to escape many local minima and even finds a correct solution in the case of the random, non-convex constraint.

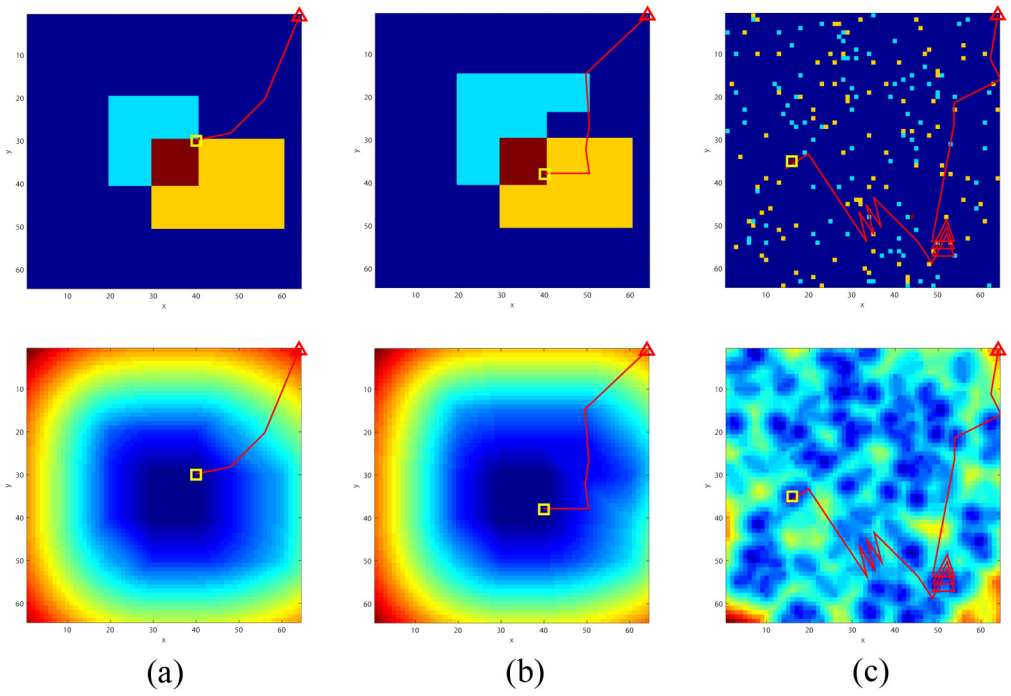

Figure 2.9 Behaviour of the DM algorithm for (a) convex constraints, (b) non-convex constraint, (c) non-convex and random constraints. Red path indicates the trajectory of the algorithm.

2.5.3.4 Variations on Basic IPAs

Martin et al. [2012] proposed a form of noise-tolerant HIO algorithm in that a tolerance with which the constraints need to be satisfied is built into the HIO algorithm. This option removes the requirement for the constraints to be satisfied exactly and allows some freedom for the IPA to explore nearby territories of the landscape.

In the context of phase retrieval, if the exact size and shape of the object is not known in advance and the only information about the support of the object is that it is finite, then IPAs can still be adapted to find the solution - along with the correct support. This is achieved by assuming an initial guess for the support and then low-pass filtering and thresholding the estimate of the support every few iterations during the execution of the IPA. Eventually, with a suitable threshold and low-pass filtering kernel, the algorithm can converge to the correct phase and the correct support. This type of algorithm is known as *Shrink Wrap* [Marchesini et al. 2003].

Alternating between two or more projection algorithms during a single phase retrieval run is also possible and has often been done since the inception of IPAs. Fienup [1982] has combined HIO with ER and Thibault [2007] mentions the combination of ER and DM as the "watched-kettle" approach. In the latter method, since the ER algorithm is effective near a solution, by carrying out a few iterations of ER for every few hundred iterations of DM, the effect is analogous to checking every-so-often to see if the kitchen kettle has boiled. The sudden change in behaviour caused by switching to another update rule causes the iterate to have a discontinuous trajectory at the point of the switch and can in general be thought of as injecting some additional "energy" into the search, allowing the iterate to be potentially pushed out of a local minimum.

Chapter 3

SERIAL FEMTOSECOND CRYSTALLOGRAPHY

3.1 INTRODUCTION

X-ray crystallography is a method for mapping out the average electron density of all the unit cells in a crystal. The density can be viewed as an image of a typical unit cell in that crystal, and at sufficiently high resolution, it can be interpreted in terms of the structure (atomic positions) of the molecule forming the crystal. By irradiating the crystalline specimen with X-rays, the coherent amplification due to Bragg scattering from all the periodically arranged molecules can in principle allow the determination of this average electron density to atomic-resolution. In practice however, various obstacles, to be elucidated in Section 3.3, prevent the imaging of important targets such as certain macromolecular proteins and viruses via X-ray crystallography. The advent of the X-ray free-electron laser (XFEL) provides a pathway around these obstacles with its ability to generate intense, coherent, and ultra-short X-ray pulses, leading to the development of the so-called *serial femtosecond crystallography* (SFX) approach [Spence et al. 2012] which has the potential to alleviate all major stumbling-blocks encountered in conventional X-ray crystallography.

In this chapter, the basic workings of an XFEL are first described, and then the barriers to structure determination of biological molecules in traditional X-ray crystallography are outlined. It is then shown that these barriers can be overcome using this new source in conjunction with the SFX method. The experimental aspects of SFX, from diffraction pattern collection to data analysis are then described, reporting on current progress and practical challenges that still await resolution.

3.2 X-RAY FREE-ELECTRON LASERS

Prior to the 1960s, all man-made sources of X-rays relied on the technology first pioneered by physicists the likes of Röntgen in the late 19^{th} century. These first devices utilised the bombardment of a metallic electrode with fast-moving electrons accelerated by an externally applied potential difference to generate X-rays. The whole contraption is enclosed in a vacated glass tube to reduce the amount of energy lost by the travelling

electrons due to collisions with background gas, hence the name given to these first generation X-ray devices: X-ray tubes. With the exception of water cooling and a spinning-electrode mechanism, both of which designed to dissipate the large amounts of unwanted heat generated by the device, no fundamental change was seen in the generation of X-rays for nearly half a century. The invention of the synchrotron in the late 1940s and their rapid development into higher energy regimes [Blewett 1998], was the first step that finally ushered in a new era of X-ray generation.

Synchrotrons produce X-rays by moving electrons inside a circular cavity where according to electrodynamics, an accelerating charge q emits electromagnetic radiation with a power P given by [Griffiths 1981]

$$P = \frac{q^2}{6\pi\epsilon_0 c^3} \frac{\dot{v}^2 - \|\mathbf{v} \times \dot{\mathbf{v}}\|^2/c^2}{(1 - v^2/c^2)^3}, \tag{3.1}$$

where $\mathbf{v}$ and $\dot{\mathbf{v}}$ are the velocity and acceleration vectors of the charge, respectively, v and $\dot{v}$ are their respective magnitudes, c is the speed of light, and ϵ_0 is the permittivity of free space. For a charged particle orbiting around a circular storage ring, its velocity vector is tangential to its trajectory whereas its acceleration vector, resulting from the centripetal force experienced by the particle as it circulates around the ring, is perpendicular to its motion. Hence, the magnitude of the cross product $\|\mathbf{v} \times \dot{\mathbf{v}}\|$ reduces to $v\dot{v}$. Inserting the expression for the centripetal acceleration, $\dot{v} = v^2/R$, into Eq. (3.1) where R is the radius of the circular storage ring, the power radiated by the circulating charge can be given in terms of the Lorentz factor, $\gamma = 1/\sqrt{1 - (v/c)^2}$, by

$$P = \frac{q^2}{6\pi\epsilon_0} \frac{v^4\gamma^4}{c^3 R^2}. \tag{3.2}$$

The instantaneous power emitted by a charge accelerating in a circular accelerator is thus limited by the minimum radius (maximum curvature) of the particle path - the other influential factor being the maximum strength of the magnetic fields that accelerates the charged particles. Second generation X-ray synchrotron sources were parasitic on so-called cyclotrons used for high energy physics experiments where the spurious, and originally unwanted, X-rays were captured and utilised. Third generation storage rings are dedicated accelerators for generating X-rays with amplification magnets called insertion devices placed along its tracks, leading to an increase in the brilliance of the generated X-rays by 12 orders of magnitude compared to X-ray tubes.

The free-electron laser in the X-ray regime, the 4$^{\text{th}}$ instalment of the extension to synchrotron technology, is outperforming synchrotrons in much the same way that synchrotrons outperformed X-ray tubes. The X-rays generated in an XFEL possess a brilliance several orders of magnitude higher than 3$^{\text{rd}}$ generation synchrotron sources. Their extremely high intensity, along with their incredible properties of a near complete spatial coherence and ability to generate femtosecond pulses - not available in any

previous sources of X-rays - make XFELs currently one of our most powerful tools for probing the detailed structures of matter.

3.2.1 Operating Principle of an XFEL

The operating principle of XFELs was established in the 1970s with Madey [1971] giving the first complete quantum mechanical description of the stimulated scattering of radiation emitted by electrons travelling through the device. Experimental work on the generation of radiation by free-electron lasers operating in the infrared and microwave region had in fact been demonstrated earlier by Motz et al. [1953], albeit the output radiation was not completely coherent. The first amplification and lasing of coherent radiation was achieved by Elias et al. [1976] and Deacon et al. [1977], respectively. Free-electron lasers for X-rays however was only able to be practically realised within recent years through the advancement of laser and accelerator technologies.

In an XFEL, the electrons are accelerated to near the speed of light and then made to pass through a series of permanent magnets that alternate in polarity in what is called an undulator. The result of this arrangement is that the electrons travel through an approximately sinusoidal magnetic field perpendicular to their direction of propagation. A particle with a charge q moving through an electric field $\mathbf{E}$ and magnetic field $\mathbf{B}$ with a velocity $\mathbf{v}$ will experience a force $\mathbf{F}$ given by

$$\mathbf{F} = q\left(\mathbf{E} + \mathbf{v} \times \mathbf{B}\right). \tag{3.3}$$

This is the so-called Lorentz force and it acts in the direction transverse to the propagation direction of the electrons. Because the magnets are placed such that the adjacent poles alternate in their polarity, the direction of the Lorentz force also alternates and thus the path that the electron bunch takes as it travels through the undulator is a sinusoidal one as depicted in Fig. 3.1.

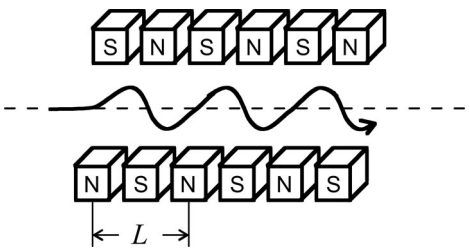

Figure 3.1 Path of an electron as it passes through an undulator. Adapted from Jackson [1999].

As mentioned above, a charge undergoing acceleration emits radiation. The emitted wavelength in this case depends on the spatial period of the oscillatory motion undergone by the charge. The period of the sinusoidal electron path is equal to the spacing between two undulator magnets, L, as shown in Fig. 3.1. However, because

the electrons are travelling near the speed of light, the spacing of the magnets from the point of view of the relativistic electrons is shrunk by the effect of length contraction to approximately L/γ. Moreover, relative to the laboratory frame, the Doppler effect further reduces this spacing to $L/2\gamma^2$.

Taking into account the transverse component of the electron bunch velocity due to the Lorentz force, the emitted radiation can be shown to have a wavelength [Saldin et al. 2000]

$$\lambda = \frac{L}{2\gamma^2}\left(1 + K^2\right), \tag{3.4}$$

where K is the so-called *undulator parameter* given by

$$K = \frac{eLB}{2\pi m_o c^2}, \tag{3.5}$$

e is the charge on an electron, B is the undulator magnetic field and m_o is the rest mass of the electron. The undulator parameter is usually relatively small compared to unity and so the emitted XFEL beam wavelength is approximately

$$\lambda \approx \frac{L}{2\gamma^2}. \tag{3.6}$$

Tunability of the emitted wavelength from an XFEL can thus be achieved through the parameter γ which varies with the energy of the accelerated electron. Injecting more energy results in a shorter wavelength. The energy needed for a specific wavelength can be approximated via the expression for the total energy of a relativistic particle, $E_T = \gamma m_o c^2$, giving

$$E_T \approx \sqrt{\frac{L}{2\lambda}} m_o c^2. \tag{3.7}$$

The energy needed to achieve 1Å radiation from a 3.4m magnet spacing (that of the LCLS undulator [Emma et al. 2010]) is then calculated via Eq. (3.7) to be around 90GeV. In reality, the exact energy requirement depends on the specific design of the XFEL and much lower energy levels than that implied by Eq. (3.7) are usually sufficient. An electron energy on the order of 10GeV is generally required to achieve 1Å wavelength radiation in practice [McNeil and Thompson 2010]. The principles of relativity thus transform a macroscopic magnet spacing into an Ångström-sized periodicity (wavelength) of X-ray radiation.

The coherence and amplification of the emitted radiation in an XFEL results from the interaction of the travelling electron bunches with the previously emitted radiation from bunches that travelled through the undulator beforehand. The amplification process starts when the first waves emitted by the oscillating electrons downstream of the undulator reach other bunches further upstream. The transverse velocity of each electron bunch, $\mathbf{v}_T$, causes it to move through the magnetic field of a previously emitted wave, $\mathbf{B}_w$, creating a second Lorentz force, $\mathbf{F}_L$, this time in the same direction as the

travelling electrons. This longitudinal Lorentz force thus pushes the electrons toward a zone of zero $\mathbf{B}_w$ as depicted in Fig. 3.2, a process referred to as *microbunching*.

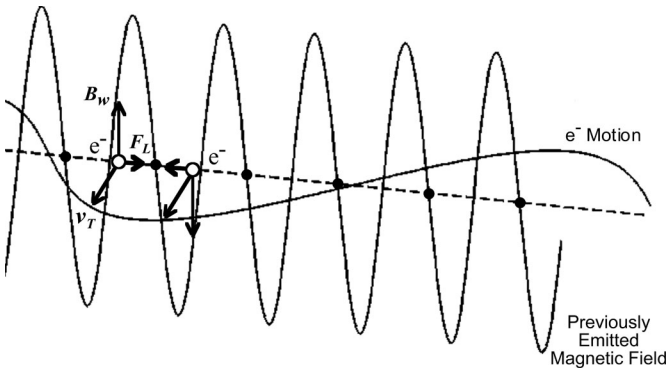

Figure 3.2 The electron microbunching process. Two electrons forming a microbunch are depicted by two white circles. The dotted line indicates the central axis of the undulator where an unperturbed electron bunch would travel and the black dots show positions where electron microbunches converge. Adapted from Ribic and Margaritondo [2012].

The unstructured electron bunch thus evolves into a series of narrow planar distributions resembling thin sheets, called *microbunches*, with the spacing between two adjacent microbunches equal to the wavelength λ of the emitted wave. The radiation emitted from the electrons in these sheets are then correlated because each microbunch is separated by exactly one wavelength after the completion of microbunching. Therefore, as opposed to the random emissions from a homogeneous electron bunch, the emissions from a set of microbunches are now all in phase, adding up coherently and reinforcing each other to cause an exponential increase in their combined intensity. The intensity increase continues until saturation is reached. Thus, the emitted waves generated by the acceleration of all microbunches amplify and superimpose coherently to give an intense output radiation of near complete spatial coherence.

The formation of amplified spatially coherent X-rays caused by the positive feedback of the electromagnetic fields of the oscillating electrons is given the name *self-amplified spontaneous emission* (SASE). The temporal coherence of the amplified beam is relatively poor however, as the SASE process starts essentially from random shot-noise. One way to achieve a smaller variation in the frequency bandwidth of the generated pulse is to use an additional undulator to produce an initial X-ray pulse and pass that through the main undulator. When an external X-ray beam is injected into the undulator along with the electron beam, the SASE process is said to have been *seeded*. This seeding procedure is desirable as it creates a more stable emission with a smaller frequency bandwidth in the resulting amplified radiation [Geloni et al. 2011].

Despite their name, free-electron lasers are not truly lasers in the sense of the word referring to conventional quantum devices. The principle of its operation make FELs more similar to that of vacuum tube devices, based on interactions of electron beams with radiation in a vacuum [Saldin et al. 2000]. Since the electrons are not confined to

atoms as in conventional material lasers, but are travelling in free-space, higher powers can be tolerated without the material restrictions imposed by a lasing medium. The operation of XFELs in the real world are much more complicated than described here. Interested readers can refer to Margaritondo and Ribic [2011] and Saldin et al. [2000] and references therein.

3.2.2 Femtosecond Time Structure

The final property of relevance to SFX experiments is the ultra-fast nature of the generated radiation pulse. The orbit of electrons in the circular track of a synchrotron dictates the duration of a radiated pulse in 3rd generation devices. The shortest pulse duration achievable for standard operating parameters on current synchrotrons is on the order of 10ps [Schoenlein et al. 2000]. Over the past few decades, a variety of approaches have been tried to generate X-ray pulses on the time scale of femtoseconds for synchrotrons [Bucksbaum and Merlin 1999]. The most successful among these is the laser slicing technique in which short X-ray pulses are "sliced" out of long synchrotron beams by modulating the energy of the electrons using an external optical femtosecond laser [Schoenlein et al. 2000]. However, all efforts to shorten the pulse length of synchrotron sources suffer from the fact that the intensity of the resultant pulse is drastically reduced during its shaping.

XFELs are immune to such a decrease in pulse intensity following on from shortened electron bunches because the SASE process re-magnifies the intensity of the reduced pulse. Short pulse lengths can thus be achieved without compromising the brightness of the output radiation. The length of an XFEL pulse is defined as the full-width-at-half-maximum (FWHM) of the electron bunch length. Pulse durations of around 10fs are currently achievable at the LCLS [Emma et al. 2010] employing the overall instrumental set-up shown in Fig. 3.3.

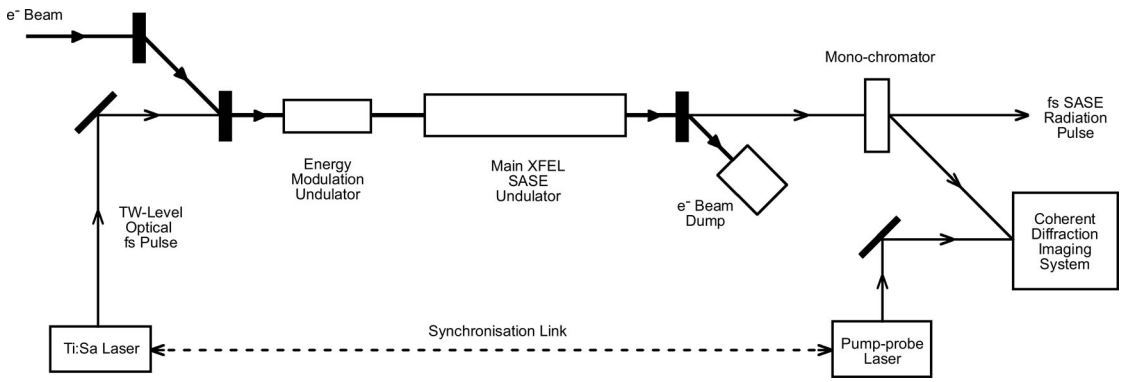

Figure 3.3 Femtosecond XFEL pulse generation schematics adapted from Saldin et al. [2004].

Short-lasting transient phenomena are typically studied with so-called *pump-probe* experiments where the target to be investigated is stimulated by a high-power pump

pulse from a laser-driven source, typically employing ultra-fast optical lasers. The excitation from the pump pulse causes the sample to transit away from its equilibrium state and the excited sample can then be probed with the main emanating radiation, in this case the XFEL beam, after a predefined time delay. Repeating this process with different delays, a time-lapsed record of the ultra-fast phenomena can be made. The extremely short duration of XFEL pulses makes them ideal for studying systems where changes of interest occur on a timescale comparable to the duration of the pulse. The XFEL beam can be synchronised with the pumping pulse via an electronic link as shown in Fig. 3.3 where delays between the two waves can be implemented using path-length differences.

3.2.3 A Survey of XFEL Facilities around the World

The Linac Coherent Light Source (LCLS) facility at the Stanford Linear Accelerator Center in USA began operation in 2009 and has since produced laser-like pulses of X-rays at around 70fs duration with approximately 10^{12} photons per pulse [Chapman 2009], and even more recently, pulses consisting of 10^{13} photons in less than 40fs pulse lengths were achieved [Boutet et al. 2012]. The LCLS is the world's first XFEL operating in the hard X-ray region ($\lambda \approx 1$Å). Before that, the Free-electron Laser in Hamburg (FLASH) facility at Deutsches Elektronen-Synchrotron (DESY) in Germany produced soft X-rays at wavelengths of around 100Å [Ackermann et al. 2007]. Other soft X-ray sources also exist, including the FERMI facility located at Trieste, Italy, and the DIpole And Multipole Output for the Nation at Daresbury in the UK, more conveniently known as the DIAMOND light source. The LCLS is currently being upgraded to a more powerful instrument called LCLS-II which will allow multiple experiments to be run at the same time. At completion, expected in 2025, the facility will operate two soft X-ray lasers and two hard X-ray lasers.

The Super Photon ring-8 (SPring-8) synchrotron located in the mountains of western Japan houses the SPring-8 Angstrom Compact Free Electron Laser (SACLA) and is the second hard X-ray XFEL in the world to come online, achieving inaugural lasing in June 2011 [Kato et al. 2012]. With a more compact design, SACLA's overall length is just 700m compared to the 3.2km accelerator arm of the LCLS.

A major upcoming facility is the European XFEL, currently under construction between the DESY site and the town of Schenefeld, Germany [Altarelli et al. 2006]. Much denser sequences of electron bunches will be able to be aligned in this XFEL to produce a pulse repetition rate as high as 30kHz. The European XFEL is scheduled for user operation in 2017. Numerous other XFELs are currently being commissioned throughout the world. Two more facilities of note are the SwissFEL at the Paul Scherrer Institute (PSI) in Switzerland and the PAL-XFEL at the Pohang Accelerator Laboratory in South Korea. The healthy interest in XFELs could be attributed to the

growing recognition of particle accelerators as a key component for not only physics experiments but also for non-material science applications such as structural biology. A list of current and near-future XFELs is made in Table 3.1 along with some of their key parameters. A more detailed but possibly now outdated list can be found in Ribic and Margaritondo [2012].

Table 3.1 Current and planned XFELs around the world.

FEL Name	Affiliated Facility	Location	Wavelength Range[a] (Å)	Pulse Duration (fs)	Max Pulse Repetition (Hz)	Status[b]
LCLS	SLAC	USA	26-1.2	10-500	120	Operational
FLASH	DESY	Germany	440-41	10-50	8×10^3	Operational
SACLA	SPring-8	Japan	610-0.6	100	60	Operational
FERMI	ELETTRA	Italy	1000-100	25-200	50	Operational
FLASH-II	DESY	Germany	800-40	10-200	8×10^3	Operational
SPARX	Tor Vergata	Italy	400-6	30-200	100	2015
POLFEL	ASI	Poland	> 270	10-100	1×10^5	2015
PAL-XFEL	PAL	Korea	50-0.6	0.5-50	60	2016
JLAMP	JLab	USA	1000-100	10-100	1×10^6	2016
Shanghai FEL	SINAP	China	> 90	100	50	2016
European XFEL	EXFEL GmbH	Germany	> 0.5		3×10^4	2017
SwissFEL	PSI	Switzerland	70-1	2-13	100	2019
LCLS-II	SLAC	USA	0.5	20	9×10^5	2021

[a] Wavelength range of the fundamental lasing harmonic.
[b] Year denotes time of expected operation.

3.3 BARRIERS TO STRUCTURE DETERMINATION

Three major difficulties have plagued crystallographers since the introduction of the technique of X-ray crystallography just over one hundred years ago by Friedrich, Knipping and von Laue [Friedrich et al. 1912]. The first difficulty is that many target molecules are difficult to crystallise, especially complex biological structures. Second, if reconstructions of the molecular density at greater resolution are desired, the incident X-ray flux must be increased, leading to a high radiation dose that damages the sample and counteracts the gain in resolution. Third, the inability of current and past X-ray detectors to register the phase of the diffracted wavefront prevents a direct inversion of the diffraction pattern via the inverse Fourier transform.

These barriers are now briefly explored, giving an account of some of the methods previously employed to deal with them, and then it is shown that these barriers can potentially be alleviated, all at once, by the XFEL source in conjunction with the SFX method.

3.3.1 Crystal Formation

The first step in X-ray crystallography is to obtain well-diffracting crystals. There is a definite component of art involved in the crystallisation of complex biological structures such as proteins due to the highly experimental nature of the procedures. Both experience and luck are necessary to produce good quality macroscopic crystals and a significant amount of time is usually required for the former to be gained and the latter to come by. The effort it takes to fine-tune the crystal production process is arduous and therefore forms a major bottleneck in molecular structural determination.

The size of a crystal is important as the diffracted intensity at the Bragg peaks is proportional to the square of the total number of unit cells in a crystal, hence the larger the crystal, the higher the signal at the Bragg reflections. However, complex molecules are difficult to isolate and grow into large, well-diffracting crystals. Membrane proteins are a prime example of such difficult-to-crystallise targets due to their hydrophobic nature and instability when removed from their native membrane environment [DeLucas 2009]. A form of crystallisation known as lipidic cubic phase (LCP) crystallisation has been successfully applied to circumvent the above problems for membrane proteins and have been shown to produce well-diffracting 3-D crystals [Landau and Rosenbusch 1996]. LCP utilises lipids as a medium for the crystals to grow, mimicking the natural lipid bilayer of biological membranes where the proteins are kept hydrated by interconnecting water channels. The protein array further organises itself such that it forms a minimal surface structure, behaving as a gel-like solid macroscopically. The LCP method can prompt membrane proteins to produce highly ordered crystals, however, these crystals are again often limited in their size.

Crystal disorder and dislocations in the internal structure, measured by the crystal's *mosaicity*, are also issues that need to be contended with. A high degree of mosaicity means large internal defects reside in the crystal which reduces the resolution obtainable from the scattered X-rays and also broadens the Bragg reflections. Crystal imperfections can be introduced by dynamic flows in the crystallising fluid and the settling of the crystal material to the bottom of the crystallisation vessel due to Earth's gravity. Thus, crystals have been trialed to grow in space [Littke and John 1984] and in microgravity environments such as on-board the space shuttle [Snell et al. 1995] in order to minimise defects and inhomogeneity seen with crystals grown terrestrially. Such experiments were successful in that they produced well-ordered, good quality crystals but at the expense of operating in extreme locations of outer space which is not viable at least in the foreseeable future. Note that some degree of mosaicity is still needed for reflections to be observed for macroscopic crystals in practice.

Using SFX, only small crystals on the nanometre scale (nanocrystals) are needed. The large number of incident photons means that a useful diffraction pattern can be obtained even without the Bragg amplification usually required from a large num-

ber of unit cells in a macrocrystal. The advantage of needing only small crystals is that nanocrystals can form spontaneously and frequently in their host crystallisation medium, referred to as their mother liquor, thus reducing the lengthy and time-consuming steps typically required to find the correct conditions for macrocrystal growth. It has also been suggested that nanocrystals may exhibit a lower degree of disorder compared to larger crystals [Cusack et al. 1998] since long-range defects are less likely to develop inside a smaller volume. Analysis from X-ray powder diffraction of hen egg white lysozyme, where the sample being diffracted is in the form of a powder rather than an intact crystal, provide some evidence to support this theory [Dreele 2007]. Note however that the small sizes of nanocrystals conversely mean that surface effects need to be taken into account as the surface-area-to-volume ratio is much larger for smaller crystals - more will be said on this in Chapter 6.

3.3.2 Radiation Damage

A delicate balancing act is always necessary when using imaging methods that alter the target under study. X-ray crystallography is no exception. Increasing the intensity of the incident beam generates stronger scattering at larger angles, giving measurements at correspondingly higher resolutions, but at the same time, the ionising effect of X-rays means that the intensity increase also boosts the probability of alterations in the atomic structure of the sample after passage of the beam. The problem of radiation damage is especially critical for biological specimens. A quantitative measure of the amount of radiation experienced by a material can be given by the amount of energy absorbed by the material from the radiation per unit mass of the material [Drenth 2007]. The SI unit for radiation dosage is the Gray (Gy) where 1Gy = 1J/kg.

At the photon energies of interest to X-ray crystallography with biological specimens, on the order of 10keV, 2% of the incident X-ray photons interact with the sample while the rest pass straight through. Of the 2% of photons that interact, only 8% generate useful scattering (elastically scattered), the rest are scattered inelastically (8%) or absorbed via the photoelectric effect (84%) [Ravelli and Garman 2006]. The initial emission of photoelectrons are followed by a cascade of lower-energy electrons due to secondary impacts and field ionisations, occurring on a timescale on the order of 10fs [Barty et al. 2012]. The ionisation due to direct photo-absorption and subsequent secondary processes affect the diffraction pattern in the following three ways [Caleman et al. 2011], listed approximately in their order of severity from the least to the most severe:

1. Ionised atoms have a smaller elastic X-ray scattering power and thus become less effective at generating useful scattering.

2. Free electrons that do not have enough energy to escape the sample remain as a

background electron gas which contributes to noise in the diffraction pattern.

3. Removal of electrons from atoms leaves behind positively charged ions that repel each other due to Coulomb forces, leading to the destruction of the sample in a so-called "Coulomb explosion."

Traditionally, to deal with radiation damage, samples are cooled to liquid nitrogen temperatures (on the orders of 100K) to reduce the diffusion of ionised particles throughout the crystal [Garman 1999]. The so-called Henderson safe-dose limit for cryogenically cooled crystals is 33MGy [Henderson 2006].

By using XFELs with their ultra-brief X-ray pulses, the issue of radiation damage becomes almost non-existent. Fig. 3.4 compares the interception of a molecule by a pulse shorter than the Coulomb explosion time with a pulse longer than this time. Since information about the molecule is essentially contained within the exiting pulse, it can be appreciated that a sluggish pulse will be caught in the "blast" and the ensuing carnage of the exploded sample thus distorting information about the undamaged molecule captured before the explosion. The effect of radiation damage will therefore not show up in a diffraction pattern if the X-ray pulse is terminated before the photoelectron cascade leading to a Coulomb explosion vaporises the sample, a method coined the "diffract-before-destroy" mode of coherent diffraction imaging [Spence and Doak 2004].

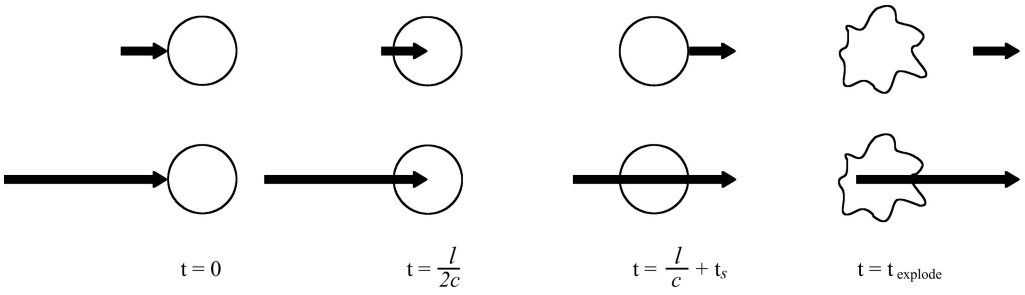

Figure 3.4 "Out-running" radiation damage using an extremely brief pulse. The arrowheads represent the X-ray pulse and the circles the molecule. l denotes the linear size of the circular molecule, c is the speed of light and t_s is the pulse duration of the shorter pulse. The molecule disintegrates after time t_{explode}.

The idea that the speedy termination of coherent radiation with a high enough incident intensity might provide a diffraction-limited image of organic materials before they are destroyed was first analysed in detail by Solem [1986]. He identified the two major factors influencing the amount of X-ray exposure needed to form a usable image: one, enough scattered photons to obtain a statistically significant signal and two, the efficiency of recording those photons. Neutze et al. [2000] then provided one of the first computer models that estimated the obtainable resolution for a given pulse duration and intensity by simulating atomic motions following the photo-ionising impact of the incident X-ray. Their simulations showed that by using a combination

of very high doses and very short exposure times, useful structural information may indeed be extracted before radiation damage destroys the sample. Experimental results that demonstrated the ability of ultra-fast pulses to avoid radiation damage were first presented by Chapman et al. [2006a] where they obtained diffraction patterns from an artificially etched sample using the FLASH soft X-ray FEL. The sample was shown to be completely destroyed after the exposures to the XFEL beam while reconstructions of the diffraction pattern via phase retrieval algorithms showed the image of the original sample without damage.

For further work on the implications of femtosecond radiation on the structure of materials, see Jurek et al. [2004] and an interesting discussion more recently by Barty et al. [2012], proposing that the diffracted X-rays can effectively be "gated" by the loss of crystallinity in the sample and thus producing an apparent pulse length shorter than the duration of the actual incident pulse.

3.3.3 The Phase Problem

Since the unit cells within a crystal are positioned periodically in space, the Fourier transform of a large crystal, i.e. the diffracted X-ray amplitudes, are discrete in Fourier space and we observe only the Bragg reflections with no scattered signal in-between. In crystallography, the value of the complex diffraction pattern at the Bragg peaks whose positions are indexed by the Miller indices $\mathbf{h}$, is referred to as the *structure factors*, denoted by $F_{\mathbf{h}}$. Each structure factor can be written in terms of a magnitude, $|F_{\mathbf{h}}|$, and a phase, $\phi_{\mathbf{h}}$, component where

$$F_{\mathbf{h}} = |F_{\mathbf{h}}| \exp\left(i\phi_{\mathbf{h}}\right). \tag{3.8}$$

The intensities $|F_{\mathbf{h}}|^2$, up to some maximum index $\mathbf{h}_{\max}$, is readily recorded by a detector in a diffraction experiment. If the phases associated with these amplitudes can also be determined, then the 3-D electron density map of the molecule, $f(\mathbf{x})$, up to a resolution determined by $\mathbf{h}_{\max}$ can be obtained via a truncated Fourier series summation,

$$f(\mathbf{x}) = \frac{1}{V} \sum_{\mathbf{h}}^{\mathbf{h}_{\max}} F_{\mathbf{h}} \exp\left(i2\pi\mathbf{h}\cdot\mathbf{x}\right), \tag{3.9}$$

where V is the volume of the unit cell.

However, direct measurement of phases is difficult because the frequencies of X-rays are on the order of 10^{18}Hz, which means an instrument capable of measuring the phases of an X-ray beam must be able to detect changes in an electric field occurring on a time scale on the order of 10^{-18}s. Such an instrument that is also robust to distortion and measurement noise is beyond the reach of our current technology hence rendering the direct measurement of X-ray phases not yet possible and giving rise to the so-called

phase problem.

Fortunately, there exists several alternative strategies for indirectly determining the phases of the scattered X-rays in conventional X-ray crystallography and coherent diffractive imaging (CDI) experiments in general:

- Molecular replacement
- Isomorphous replacement
- Anomalous scattering
- Direct methods
- Iterative reconstruction of an oversampled intensity

It is relevant to discuss both crystallographic and coherent diffractive imaging methods together in the context of nanocrystallography as CDI is usually applied to the imaging of single noncrystalline particles - i.e. the limiting case of a sufficiently small nanocrystal.

3.3.3.1 Molecular Replacement

Two physically similar objects must also have similar Fourier transforms as Euclidean distance in real space and Fourier space is preserved. This simple observation is utilised to solve the phase problem for molecules where the structure of another, similar, molecule is already known. Solving the phase problem using knowledge of a related molecule is the idea behind *molecular replacement*. An example is when the structure of a member of a certain protein family has been determined, the other members of the same family can often be analysed by molecular replacement. The majority of the protein structures solved to date have been achieved by this method [Sherwood and Cooper 2011].

In molecular replacement, the target molecule being analysed is compared with a similar molecule of a known structure, referred to as a "search model" or "search molecule." The method then proceeds by determining the orientation and position of the search molecule relative to the unit cell of the target molecule using rotation and translation operations. After optimising for the orientation and the positions, the phases are then calculated from the search model and used in conjunction with the measured magnitudes of the target molecule to form an estimate of the electron density map. The map may then be refined by incorporating known constraints, a procedure referred to generally as *density modification* [Millane 1990, Millane and Lo 2013]. The search and target molecules must have a reasonable amount of 3-D structural agreement. As a result of the correlation between the amino acid sequence and the 3-D

structure of a protein, at least 25% of the amino acid sequence must be identical for there to be a good chance of success [Dodson 2008].

Molecular replacement is simple in principle, relatively straightforward to apply, and works well for many macromolecules. However this method fails when a completely new structure needs to be determined and it also suffers from the undesirable trait that model bias may result as Fourier phases usually contain more structural information than Fourier magnitudes, a phenomenon known as "phase dominance" [Millane and Hsiao 2009]. The calculated target structure may thus be biased towards the phase of the search molecule. Moreover, since the calculated structures are being reused in other molecular replacement sessions, errors can propagate, in a manner analogous to passing down "bad genes" in inheritance.

3.3.3.2 Isomorphous Replacement

The method of isomorphous replacement is an experimental technique for determining phase information first used in the 1950s [Green et al. 1954]. Heavy atoms are absorbed into the crystal which situate themselves identically in each unit cell, giving rise to a "heavy atom derivative" crystal. Diffraction measurements are made from the original (native) crystals and also from the heavy atom derivatives. The electron-rich heavy atoms scatter X-rays strongly and perturb the diffraction in a controlled manner which can be used to estimate the phase.

More precisely, the diffraction magnitudes resulting from the heavy atom derivative crystal, $F_{\mathbf{h}}^{D}$, become the sum of the structure factor from the native crystal, $F_{\mathbf{h}}$, and that from the heavy atoms themselves, $F_{\mathbf{h}}^{H}$, giving

$$F_{\mathbf{h}}^{D} = F_{\mathbf{h}} + F_{\mathbf{h}}^{H}. \tag{3.10}$$

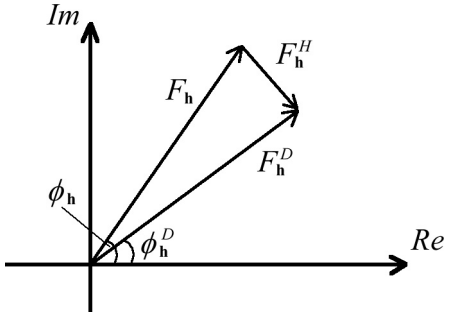

Figure 3.5 Complex structure factors and their relationships on the complex plane after a heavy atom addition. This type of plot is known in crystallography as a Harker diagram.

From Fig. 3.5, the desired phases of the native structure factors, $\phi_{\mathbf{h}}$, can be related

to the phase of the heavy atom derivatives, $\phi_{\mathbf{h}}^H$, and the three magnitudes by

$$\phi_{\mathbf{h}} = \phi_{\mathbf{h}}^H + \cos^{-1}\left(\frac{|F_{\mathbf{h}}^D|^2 - |F_{\mathbf{h}}|^2 - |F_{\mathbf{h}}^H|^2}{2|F_{\mathbf{h}}||F_{\mathbf{h}}^H|}\right). \tag{3.11}$$

The amount by which the complex amplitude of a given reflection changes upon adding the heavy atom, $|F_{\mathbf{h}}^H| = |F_{\mathbf{h}}^D - F_{\mathbf{h}}|$, is known as the *isomorphous difference*. The phases $\phi_{\mathbf{h}}^H$ can be determined through the so-called isomorphous difference Patterson map, $P_H(\mathbf{x})$, which is the autocorrelation of the heavy atom positions given by

$$P_H(\mathbf{x}) = \sum_{\mathbf{h}} |F_{\mathbf{h}}^H|^2 \exp\left(i2\pi\mathbf{h}\cdot\mathbf{x}\right). \tag{3.12}$$

In practice, $|F_{\mathbf{h}}^H|^2$ is not measurable and the quantity $\left(|F_{\mathbf{h}}^D| - |F_{\mathbf{h}}|\right)^2$ is used instead. The substitution is only correct if the vectors $F_{\mathbf{h}}$ and $F_{\mathbf{h}}^H$ are collinear, but nevertheless forms a good estimate of $|F_{\mathbf{h}}^H|^2$ under most circumstances [Millane 1990]. The difference Patterson map can then be approximated as

$$P_H(\mathbf{x}) \approx \sum_{\mathbf{h}} \left(|F_{\mathbf{h}}^D| - |F_{\mathbf{h}}|\right)^2 \exp\left(i2\pi\mathbf{h}\cdot\mathbf{x}\right) \tag{3.13}$$

and is dominated by the effect of the heavy atoms. Thus, $P_H(\mathbf{x})$ can be used to determine the heavy atom positions from the assumption that the largest non-origin peaks on the Patterson map indicate the heavy atom vectors, and in turn, an estimate of $F_{\mathbf{h}}$ can be obtained which leads to an estimate of $\phi_{\mathbf{h}}^H$.

A twofold ambiguity for the value of $\phi_{\mathbf{h}}$ calculated via Eq. (3.11) in the range $[0, 2\pi]$ exists due to the multivaluedness of the inverse cosine function. This can be visualised by constructing phase circles as shown in Fig. 3.6 which depicts all possible phases a particular structure factor with a given magnitude can have. The phase circle for $F_{\mathbf{h}}$ (black) intersects the phase circle for a single heavy atom derivative (blue or red) at two points. A unique solution may be found by using two different heavy atom derivatives and choosing the phase solution that coincide. In practice, the method of using a number of different isomorphous derivatives to resolve the uniqueness problem and to combat the effects of noise is called *multiple isomorphous replacement* (MIR).

The term "isomorphous" refers to the notion that addition of heavy atoms does not alter the packing and conformation of the molecules within the native crystal. However, this can be difficult to achieve in practice and one of the main sources of error is the lack of perfect isomorphism. A weighting scheme is used in isomorphous replacement that minimises the mean squared error in the electron density map due to errors in the phase angles. By assuming a normal distribution for the errors in the lengths and angles of the vectors on the Harker diagram, a probability distribution on the molecular phase angle can be derived. The best value of $F_{\mathbf{h}}$ is obtained by taking the weighted

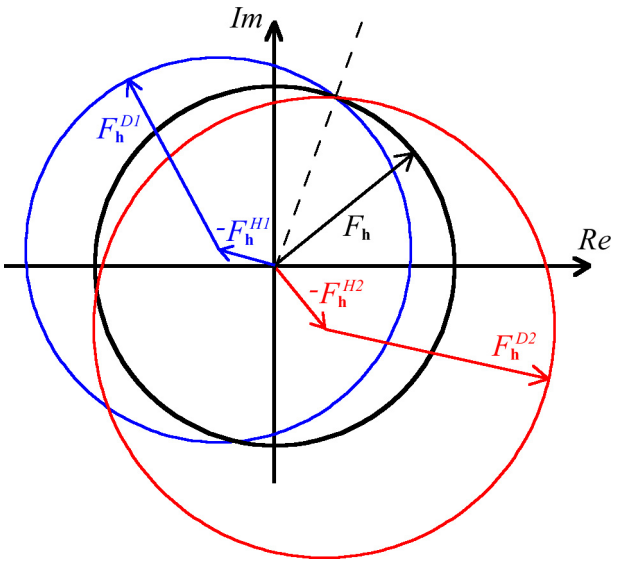

Figure 3.6 Phase circle construction for two heavy atom derivatives. The correct phase is given by the angle of the dashed line.

average over the range of possible $F_{\mathbf{h}}$ according to their probabilities. The weights so obtained are called *figures of merit* which ranges from 0 to 1 and is approximately equal to the cosine of the phase error for that reflection. The best structure factors along with their corresponding phases after the weighting procedure are then directly inverted via Eq. (3.9) to generate the best estimate of the electron density map.

3.3.3.3 Anomalous Scattering

Another important experimental phasing technique is *multiple-wavelength anomalous dispersion* (MAD). MAD is a variant of isomorphous replacement that has now surpassed MIR as the dominant phasing method in X-ray crystallography [Lattman and Loll 2008]. Its underlying phasing principle is the same as MIR except now instead of perturbing the crystal structure by the addition of heavy atoms, the perturbation is introduced through a change of incident X-ray wavelength.

The strength and phase of the scattered radiation from an atom changes as the wavelength of the X-rays approach a value near that atom's so-called absorption edge where an electronic transition occurs due to resonant effects. These effects result in the breakdown of Friedel's law, i.e. $|F_{\mathbf{h}}| \neq |F_{-\mathbf{h}}|$ which can be viewed as structure factor amplitudes from two different heavy atom derivatives. An anomalous difference Patterson map can be set up using the anomalous difference $|F_{\mathbf{h}} - F_{-\mathbf{h}}| \approx |F_{\mathbf{h}}| - |F_{-\mathbf{h}}|$ in the same spirit as the isomorphous difference Patterson map to deduce the position of the anomalous scatters. A twofold ambiguity still applies even though the data resembles those from two isomorphous replacements because the structure factors from the native crystal is not available. Therefore measurements at a number of other

wavelengths are taken to resolve the ambiguity, hence the name multiple-wavelength anomalous dispersion.

The most common anomalous scatters currently in use are selenium atoms. Selenium is not usually found in proteins but a naturally occurring amino acid containing selenium, called selenomethionine, can be incorporated into proteins in place of the amino acid methionine - an essential amino acid in many proteins [Lehninger et al. 2013].

As anomalous dispersion techniques require the wavelength of the incident X-ray to be varied, this type of phasing method only came of age when synchrotrons with tunable wavelengths were introduced. The perturbation on the diffraction are smaller in size compared to the addition of heavy atoms, but a major advantage is that the structure of the crystal remains unchanged therefore data measured at different wavelengths can be thought of as arising from perfectly isomorphous derivatives.

Anomalous scattering and isomorphous replacement methods can be combined. Many of the heavy atoms used for isomorphous replacement have strong anomalous scattering signals, hence the data from a single derivative using two different techniques can be combined to yield the phase information for the native crystal. The application of these two methods together is known as single isomorphous replacement with anomalous scattering (SIRAS).

3.3.3.4 Direct Methods

So-called direct methods in crystallography solve the phase problem by utilising the knowledge that the molecule consists of a small number of distinct atoms, referred to as the property of *atomicity*. A probabilistic construction is then applied, along with a series of relationships derived from the assumed atomicity, to estimate the phases. The set of possible phases are further reduced by imposing constraints on the electron density function $f(\mathbf{x})$, such as $f(\mathbf{x}) \geq 0$. First conceived by Harker and Kasper [1948], direct methods have grown to become a powerful collection of techniques capable of solving structures of molecules with fewer than around one thousand non-hydrogen atoms [Uson and Sheldrick 1999], provided that diffraction data of sufficiently high resolution are acquired. See Hauptman [1991], Giacovazzo [1999] and references therein for details on the methods.

Direct methods become ineffective for large molecules as the width of the probability distribution for the phases increases with the number of atoms in the structure and the distribution becomes essentially flat for large molecules. Moreover, high resolution data become more difficult to obtain for molecules that are large. However, even if the molecule is too large for direct methods to be applied, this technique can still be used to determine the positions of the heavy atoms required for isomorphous replacement and anomalous dispersion methods mentioned above.

3.3.3.5 Iterative Phasing of an Oversampled Intensity

It was first realised some 60 years ago by Sayre [1952] that if samples of the diffracted intensity could be obtained *between*, as well as *at*, the Bragg reflections, then the phase problem could be soluble from the intensity data alone. The notion of obtaining diffraction intensities between the Bragg peaks is known as *oversampling* in the literature, where the extra sampling is with respect to the Bragg samples located at $\mathbf{u} = \mathbf{u_h}$. If the sampling is to be kept uniform, then the first level of oversampling is when the sample spacing is reduced by a factor of two in each direction with respect to the Bragg peaks. The samples are then said to be located at *half-integral* Bragg spacings, given by

$$\mathbf{u} = \mathbf{u}_{\mathbf{h}/2} = \left[\left(\frac{2h_1 - 1}{2} \right) \mathbf{b}_1 + \left(\frac{2h_2 - 1}{2} \right) \mathbf{b}_2 + \left(\frac{2h_3 - 1}{2} \right) \mathbf{b}_3 \right]^T, \tag{3.14}$$

where h_i are integers and $\mathbf{b}_i$ are the reciprocal lattice basis vectors as described in Sec. 2.3.5 for $i = 1, 2, 3$. Since the diffracted intensity corresponds to the Fourier intensity of the unit cell, and the spectrum of that intensity is the autocorrelation of the unit cell, which is at most twice as large as the unit cell in each direction, sampling the diffraction pattern at $\mathbf{u} = \mathbf{u}_{\mathbf{h}/2}$ therefore uniquely determines the diffracted intensity by the sampling theorem.

Once the Fourier intensity is determined, the problem is then to find the electron density that gave rise to that intensity. Iterative methods such as IPAs described in Sec. 2.5 can be applied to recover the electron density of the molecule. The formulation is that of trying to find the intersection between two constraint sets: (1) the set of all possible functions that have the measured Fourier modulus and (2) the set of all possible functions that are contained within the finite support of the unit cell. This is the phasing method that is central to the investigations of this thesis. Uniqueness of this inverse problem is important and is considered below in more detail.

It is important to realise that the process of oversampling is solely for the unique determination of the object's Fourier intensity. Increasing the oversampling beyond that which is required by the sampling theorem does not add any new information to the problem. The uniqueness of the phase problem comes down to the support of the object and its autocorrelation.

Elser and Millane [2008] showed that uniqueness can be described using the so-called *constraint ratio*, denoted by Ω, defined in terms of general constraint satisfaction problems as the ratio of the number of known independent data to the number of unknown free parameters in the problem. For crystallographic phase retrieval from an oversampled intensity, the number of free parameters in the unit cell electron density $f(\mathbf{x})$ is proportional to the number of samples in the unit cell at a particular resolution. Since the electron density of the unit cell has finite support, its autocorrelation $A(\mathbf{x})$ also has finite support, and the number of independent data is proportional to half the

number of samples in the autcorrelation. The factor of a half is due to the autocorrelation being always conjugate centrosymmetric, i.e. $A(\mathbf{x}) = A^*(-\mathbf{x})$, and therefore only half of the samples are independent. The number of samples in any object is equivalent to the number of samples in its support, and so denoting the support of $f(\mathbf{x})$ by $\mathcal{S}$ and the support of $A(\mathbf{x})$ by $\mathcal{A}$, the constraint ratio is given by

$$\Omega = \frac{|\mathcal{A}|}{2|\mathcal{S}|} , \tag{3.15}$$

where the operation $|\cdot|$ gives the number of samples in the entity of the operand. Uniqueness of the phase problem requires that the number of independent data exceeds the number of free parameters in the object, which corresponds to $\Omega > 1$. The case $\Omega = 1$ is marginal in the sense that it may admit a one-parameter family of solutions, and some additional information about the problem may be required to restore uniqueness.

Objects with supports that are convex and centrosymmetric give the smallest Ω in any given dimension, with $\Omega \geq 2$ in 2-D and $\Omega \geq 4$ in 3-D. The constraint ratio is bounded below in general by

$$\Omega \geq 2^{D-1}, \tag{3.16}$$

where D is the dimension of the object, and is equivalent to saying that $|\mathcal{A}| \geq 2^D|\mathcal{S}|$. For non-convex supports, Ω can take on any value that satisfies the bound given by Eq. (3.16). For example, for a 2-D triangular support, $\Omega = 3$ and for a support with a hole, Ω can be made as large as desired by varying the size of the hole with $\Omega \to \infty$ for a support with an infinitely thin shell.

Note that even when $\Omega \geq 1$, the phase problem may still not be unique. The uniqueness properties for phase retrieval have been investigated by Bates [1982] who provided one of the earliest theoretical analysis on this matter. He defined the conditions required for different signals to have the same autocorrelation and labelled these functions as having the same "image-form." The signals listed below have the same autocorrelation as the signal $f(\mathbf{x})$

- $f(\mathbf{x} - \mathbf{a}) \exp(ib)$
- $f^*(-\mathbf{x} - \mathbf{a}) \exp(ib)$

where $\mathbf{a}$ is a constant real vector and b is a real number. Both signals above have the same Fourier intensity and therefore the same autocorrelation as $f(\mathbf{x})$. The ambiguity corresponds physically to an arbitrary shift and/or an inversion (referred to sometimes as a "twin"), and/or a phase shift which only affects the object if it is complex. Note further that an infinite number of other ambiguities may exist if $f(\mathbf{x})$ can be expressed as a convolution of two other functions. Bruck and Sodin [1979] and Barakat and Newsam [1984] have also shown using the Z-transform and polynomial factoring that except for pathological cases, the phase problem is unique in two or more dimensions.

The uniqueness of the phase problem shall be investigated further in Chapter 7 when the retrieval of phases of multiple objects from their averaged Fourier intensity is considered.

In summary, the solution of the phase problem in crystallography using a finely sampled diffracted intensity can essentially be broken down into two separate steps, the first is to obtain a correctly sampled Fourier intensity to allow for a unique autocorrelation function, and the second is to find the object corresponding to that Fourier intensity/autocorrelation. The object reconstructed in this method is the electron density of the unit cell. The process of oversampling is associated with the uniqueness of the first step, i.e. to uniquely determine the diffracted intensity via the sampling theorem, whereas the uniqueness of the object to be reconstructed from that Fourier intensity, associated with the second step, is determined by the constraint ratio Ω.

Note finally that Ω can actually still be defined for a crystalline object, i.e. in the context of phase retrieval without oversampling and using only the Fourier intensity at the Bragg positions. Solving the phase problem in this scenario is permissible if the amount of known density in the unit cell is greater than 50% [Millane and Lo 2013], which is the case for many protein crystals with high-solvent contents. With non-crystallographic symmetry, the amount of known density/solvent content needed can be reduced by a factor equal to the order of the space group of the non-crystallographic symmetry at hand.

In terms of the projection algorithm used in the second step of the reconstruction from an oversampled intensity, the two projections required for the application of IPAs can naturally be separated into one corresponding to constraints in real space and the other to constraints in Fourier space. The two basic projections are described below.

Let $\mathbf{x} = [x_1, \cdots, x_n]^T$ be the sample values of an n-sample array that contains the electron density of the unit cell. The support constraint requires the electron density to be zero outside a support region $\mathcal{S}$. The projection operator, P_S, that achieves this is one that sets all values of the image outside the support to zero, and all values of the image inside the support are left unchanged. So for all $i = 1, \cdots, n$,

$$P_S x_i = \begin{cases} x_i & i \in \mathcal{S} \\ 0 & \text{otherwise.} \end{cases} \tag{3.17}$$

Figure 3.7(a) illustrates this projection with an image that has three pixels. The support is defined to be pixels 1 and 2 and so the projection P_S will set the value on pixel 3 to zero, corresponding geometrically to a vertical projection of the point directly onto the support plane traced out by all possible values of pixels 1 and 2, hence validating intuitively that P_S can indeed be thought of as a projection.

Similarly, the projection P_M sets the magnitude of a complex number to the desired magnitude whilst leaving the phase unchanged. The set of all complex numbers that

have the same magnitude defines a circle on the complex plane, therefore the projection involves moving the target point radially onto the closest point on the circle, as can be seen in Fig. 3.7(b). The Fourier transform and the inverse Fourier transform operation required to move the point back-and-forth between the object domain and the reciprocal space domain are incorporated into the P_M projection operator itself; this is valid as Euclidean distance is preserved under the operation of a Fourier transform as mentioned before. Letting the complex vector $\mathbf{u} = [u_1, \cdots, u_n]^T$ be the Fourier transform of $\mathbf{x}$, such that $\mathbf{u} = \mathcal{F}\{\mathbf{x}\}$, the magnitude projection operator is defined as

$$P_M x_i = \mathcal{F}^{-1} \begin{cases} \dfrac{|u_{Mi}|}{|u_i|} u_i & i \in \mathcal{M} \\ u_i & \text{otherwise,} \end{cases} \tag{3.18}$$

where $\mathcal{M}$ is the set of all Fourier magnitude samples available for use in the phase retrieval problem and u_{Mi} is an element of the set of measured magnitudes that corresponds to u_i. The magnitude constraint is non-convex because the geometry of the constraint surface in this case is the intersection of $|\mathcal{M}|$ $(2n-2)$-dimensional hypercylinders in a $2n$-dimensional space, where $|\mathcal{M}|$ is the total number of Fourier magnitude samples available.

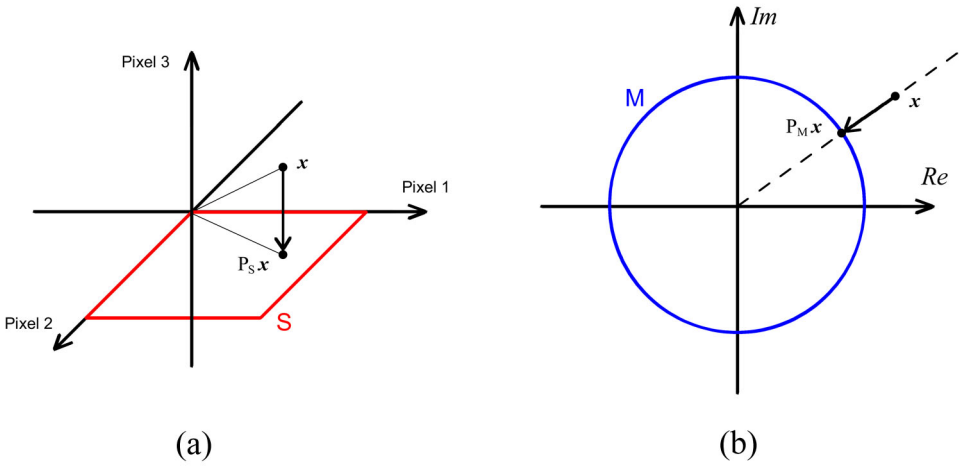

Figure 3.7 Geometrical visualisations of (a) the support projection for a three pixel image and (b) the magnitude projection for a single pixel in phase retrieval.

The application of iterative methods to X-ray imaging grew out of attempts at phasing single, non-periodic objects from their diffracted intensities - the goal of coherent diffractive imaging. CDI was investigated initially with experiments at the Brookhaven synchrotron [Sayre et al. 1998] where images of lithographed characters were reconstructed from soft X-ray diffraction patterns [Miao et al. 1999]. More recently, high resolution 3-D reconstructions of a fabricated pyramidal structure composing of gold balls were achieved by Chapman et al. [2006b] without imposing any

prior knowledge on the sample using the Shrink Wrap algorithm. Reconstructions of biological samples using the CDI technique have proved to be more difficult but increasing successes have been reported, most notable include the first 2-D [Seibert et al. 2011] and then 3-D [Ekeberg et al. 2015] images of a giant virus recovered from data collected at an XFEL. Comprehensive reviews and historical accounts of coherent diffractive imaging are given by Spence [2008] and Thibault and Elser [2010].

One reason for the relatively recent emergence of iterative computational phasing methods using a finely sampled diffraction pattern is that the sampling and the iterations require an enormous amount of processing on large arrays of numbers. Reconstruction from say a $1024 \times 1024 \times 1024$ data set can take a significant amount of time even with modern computers and the fast-Fourier transform. However, with the notable successes of the works described above and advances in computing hardware and algorithms, iterative reconstruction methods are now playing a crucial role in providing the solution to the phase problem.

3.3.4 Breaking Down the Barriers with XFELs

It is now clear that the fundamental limitation to imaging given by the issue of radiation damage can be bypassed using XFELs, leaving experimental issues such as the coherence of the pulse and the repetition rate of the XFEL as the new factors that limit the structure determination process in femtosecond nanocrystallography. If these experimental problems can be solved, a method of phasing non-identical, randomly oriented nanocrystal diffraction patterns would allow XFELs to address the three most important problems of X-ray crystallography described in the sections above: crystal quality, radiation damage and the phase problem.

The advantages of structure determination using femtosecond XFEL pulses are summarised here:

- Small crystals can form spontaneously and more frequently in their host solutions and may also exhibit a lower disorder.
- Radiation damage can be reduced or avoided altogether.
- No extreme cooling is needed to curb radiation damage, making room temperature experiments viable where proteins are in a hydrated environment so that their natural functions can be observed.
- Additional methods for solution to the phase problem are potentially available.
- Time-resolved studies at the femtosecond scale become possible.
- Highly automated structure determination processes can be achieved. For example, long data collection times using hand-mounted crystal samples are no longer required (elaborated in the next section).

3.4 DIFFRACTION DATA COLLECTION

Up until the invention of the XFEL, the experimental data collection process in X-ray crystallography has remained fundamentally unchanged since its inception over one hundred years ago. The process involves placing a macroscopic crystal on a mechanical rotation device called the goniometer, and collecting a set of diffraction patterns by rotating the crystal to various orientations relative to the X-ray beam. The nature of the XFEL source means that this conventional approach is no longer viable since the crystal will be destroyed within one single XFEL pulse, and thus giving rise to the necessity to develop new ways of crystallographic data collection, culminating in the SFX method.

The overall data collection set-up is depicted in Fig. 3.8 and involves streaming prepared samples into the path of the XFEL pulse in a vacated chamber. Diffracted signal from the interaction between the sample stream and the pulse is then recorded with a detector that has a central opening to allow the undiffracted pulse to pass through. Each of these components are described in detail in this section.

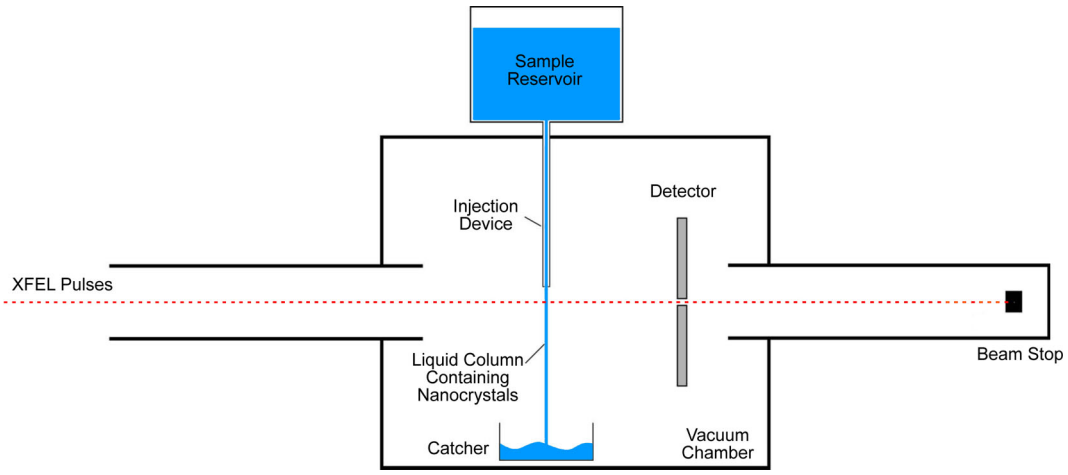

Figure 3.8 Pictorial view of an archetypical SFX experiment showing the crucial components; adapted from White et al. [2013].

3.4.1 Sample Preparation

The frequent observation of "showers" of microcrystals in their mother liquor [Georgieva et al. 2007] means that the crystallising solution itself can be utilised directly to provide a convenient medium for holding the nanocrystals needed for SFX. The maximum size of the crystals within the solution can be restricted using filters with different mean pore sizes depending on the need of the particular experiment.

An interesting problem that can arise is that although the crystals are filtered to limit their maximum size before the experiment, crystals waiting to be injected into

the beam remain bathed in their host solution and would allow regrowth of larger crystals after filtration. The effect of this so-called "recrystallisation problem" can be lessened by (1) minimising the salt content of the solution, (2) decreasing the time difference from filtering to delivery, or (3) by slowly stirring the sample solution during the experiment [Lomb et al. 2012].

3.4.2 Sample Delivery

A number of strategies for presenting the sample of interest to the XFEL beam have been developed, these include:

- Aerosol injector
- Gas dynamic virtual nozzle
- Lipidic cubic phase injection
- Electrospinning
- Fix-targets

All methods except the last listed above involve the flow of a stream of samples in a serial fashion and focusing the flow before it enters the X-ray interaction region. The focusing is usually done by increasing the velocity of the stream to reduce its cross-sectional area.

The first type of injector developed is an aerosol spray mechanism similar to that used in mass spectrometry, specialising in delivering single particle samples to the XFEL beam [Bogan et al. 2010]. Sample solutions are first atomised from an electrospray device, inducing charges in the resulting aerosol particulates. Several electrostatic lenses arranged in a stack-like configuration then filter and focus the atomised sample solution. The aerosol injection method is able to reduce the noise associated with spurious diffraction of the water envelope surrounding the specimen, referred to as the "water background," since the enveloping liquid evaporates when the particle flies through the injector. For proteins, the existence of a water jacket is required for them to be maintained in their native, functional state, thus protein samples are not suitable for delivery with the gas-phase aerosol injector. Viruses on the other hand are relatively insensitive to their surrounding environment compared to proteins and are thus able to be delivered using the aerosol injection scheme [Seibert et al. 2011].

A second type of injector is the liquid microjet injection apparatus [Weierstall et al. 2012] that is able to introduce fully hydrated nanocrystals into the pulsing XFEL beam. A hollow glass capillary tube is placed inside another one of a smaller radius, forming an inner and outer section. The sample solution is pumped through the inner capillary, and a high pressure gas, typically helium, through the outer sheath. At the

exit of the injector, the sample solution then passes into the reaction chamber under the influence of the high pressure coaxial gas, with the gas flow focusing and stabilising the sample stream. This setup, shown in Fig. 3.9, is known as the gas dynamic virtual nozzle (GDVN) and has been the workhorse sample delivery method since its first use in the earliest SFX experiments. Because the chamber where the XFEL beam intercepts the sample stream is evacuated in order to minimise the attenuation of the XFEL pulses and to reduce scattering by background molecules, the sample stream entering the high vacuum undergoes a cooling effect due to evaporative thermodynamic processes [DePonte et al. 2008]. Thus, another responsibility of the coaxial gas flow is to surround the liquid jet and prevent it from freezing in vacuum.

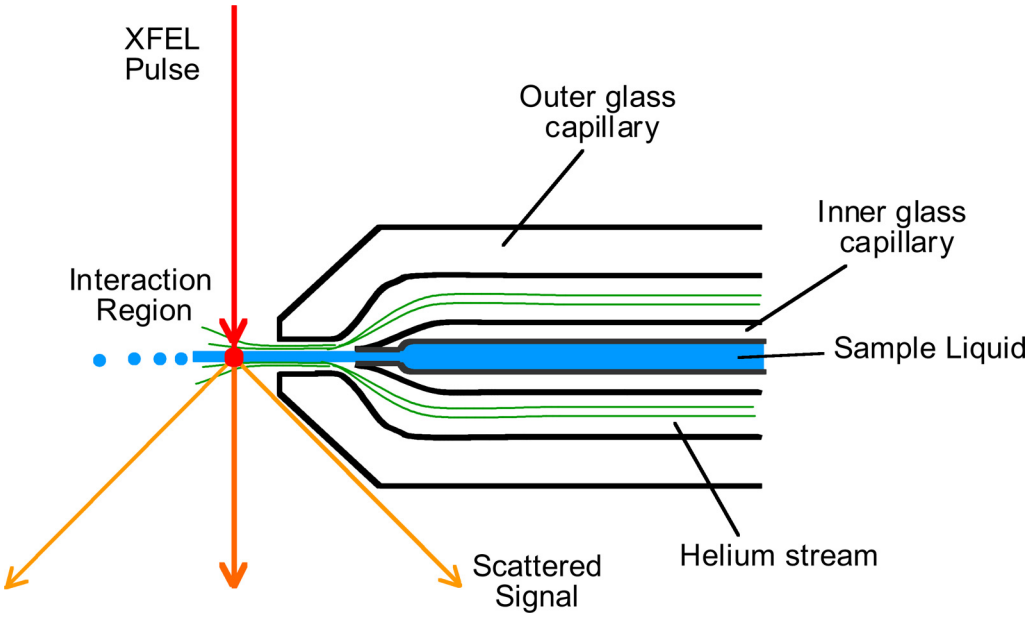

Figure 3.9 The gas dynamic virtual nozzle method of sample injection.

For a free-running liquid or gas phase jet across the XFEL beam, samples are wasted during the time interval between each XFEL pulse. This wastage becomes significant for experiments that run continuously on the order of days. The mismatch between flow rate of the GDVN method and the XFEL pulse rate results in a need for large amounts of sample, which is often not viable. A variant of the GDVN delivery scheme coined the "drop-on-demand" (DoD) method has been developed to reduce sample wastage where each droplet is generated with a piezo actuator synchronised with the XFEL pulse [Weierstall et al. 2008]. However, droplets generated by the DoD method are currently still too large to be used with single particles or even nanocrystals. More recently, work done by Abdallah et al. [2013] on integrating sample delivery techniques with microfluidic devices that are able to generate finer and more accurately controllable droplets, may allow the drop-on-demand method to be finally realised.

Another way to curb the problem of sample wastage in the interval of the XFEL pulse is delivering the samples through the LCP state described in Sec. 3.3.1. The so-called sponge lipidic cubic phase [Caffrey et al. 2014] has recently been shown to be a convenient medium for delivering microcrystals of membrane proteins to an XFEL beam. Because of its semisolid nature, the LCP jet does not readily form a free-flowing column of liquid but behaves like a viscous gel-like out-pour, similar to tooth paste. The crystal flow rate can be adjusted to match the repetition rate of the XFEL pulse so that sample consumption can be minimised. The amount of protein needed for experiments can be significantly reduced in this way [Weierstall et al. 2014, Caffrey et al. 2014]. Avenues to allow the LCP jet method to be adapted to proteins other than the membrane types have been explored [Aherne et al. 2012] and are currently being fervently pursued.

Two other currently less-widely used methods of sample delivery are the so-called fix-target and electrospinning methods. The fix-target method deposits protein microcrystals on an array of ultra-thin silicon nitride membrane before aligning the array to the XFEL beam [Hunter et al. 2014]. The electrospinning technique creates an elongated stream of free-flowing samples and is able to control the flow rate of the stream via the application of an electric field [Sierra et al. 2012]. Both methods are again designed to reduce sample consumption. However, samples delivered using the fix-target method will need to be protected from dehydration (e.g. by embedding them in oil) [Hunter et al. 2014] and the data acquisition rate is likely to be limited to a few tens of patterns per second due to the time and precision required for mechanically moving the array. A disadvantage of the electrospun microjet is its sensitivity to freezing when using crystal solutions with high water contents. Also, concerns exist regarding the potential impact of the electric field on the biological sample in the stream. Thus, the GDVN system is still currently the most widely used sample delivery technique for SFX experiments.

3.4.3 Interaction with the XFEL Pulse

Considering here the GDVN sample delivery scheme with a liquid jet. Positioning of the liquid stream and in turn the interaction region with the XFEL pulse is influenced by two factors: (1) the temperature drop as the sample jet passes into vacuum, and (2) the break up of the jet downstream due to Rayleigh instability [DePonte et al. 2008].

As mentioned in Sec. 3.4.2, a narrow droplet stream injected into a vacuum will rapidly cool due to the effect of evaporative cooling. It has been shown that for the parameters of the liquid jet injector at LCLS, a cooling rate of up to $10^6 \mathrm{Ks}^{-1}$ is experienced by the sample jet as it streams into the evacuated chamber [Spence et al. 2012]. The problems posed by this cooling effect can largely be bypassed by intercepting the sample stream with the XFEL beam close to the delivery nozzle opening before the

temperature drop has become significant. Thus, the tip of the nozzle is tapered at the opening to prevent blocking of the scattered signal, as shown in Fig. 3.9. For the experiments undertaken so far at LCLS, the interaction region is situated at about 100μm from the nozzle tip with the associated temperature drop calculated to be less than 10K [Spence et al. 2012]. If ice forms then the XFEL pulse has to be stopped immediately to protect the detector from damage due to the strongly diffracting ice.

The second factor that influences the placement of the interaction region is a phenomenon known as the Plateau-Rayleigh instability [Rayleigh 1879] which causes a continuous liquid stream to break up into smaller droplets by virtue of the tendency of liquids to minimise their surface area. Thus it is desirable to position the XFEL beam before this breakup occurs to eliminate artefacts of the diffraction due to interaction with the shape of the broken jet.

The relative sizes of the XFEL pulse and the liquid jet is shown in Fig. 3.10. The diameter of the sample jet column is typically around 3μm and the XFEL pulse has a diameter of around 0.3μm. The size of the nanocrystal varies depending on the particular biological sample in question but is typically on the orders of 0.1μm. The velocity of the sample flow in a GDVN system is around 10ms^{-1} [DePonte et al. 2008]. At the LCLS, an in-vacuum microscope camera is positioned near the injector nozzle that can be used to determine the velocity of the jet and is useful for the mechanical alignment of the nozzle. The camera helps the operators to monitor for ice and the state of the experiment in general.

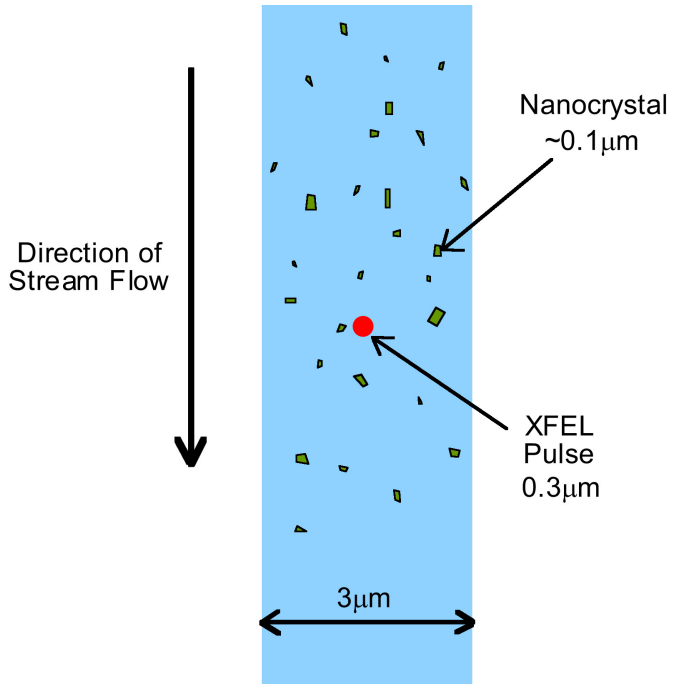

Figure 3.10 Respective dimensions of the liquid delivery jet (blue column), nanocrystal samples (green shapes), and the XFEL beam (central red dot).

The *hit-fraction* (not to be confused with the hit-rate) is the fraction of the total XFEL pulses that intersected a single sample in the delivery stream. The hit-fraction varies depending on: (1) the concentration of the sample, (2) the diameter of the X-ray beam, (3) properties of the delivery jet - in particular its size, flow rate and stability. A typical value of the hit-fraction for an SFX experiment is around 10% of the total number of patterns obtained. The other 90% of the patterns contain either multiple crystals or no crystals with the XFEL pulse only interacting with the sample fluid.

The *hit-rate* is defined as hits per unit time which is coupled to the *sample delivery efficiency* defined as hits per unit number of particles injected into the beam. The hit-rate and sample delivery efficiency determine how long the experiment will take, whilst the hit-fraction determines how much sample is needed regardless of time.

It is speculated that a spacing of around 50μm on the sample jet is needed between the XFEL pulses to avoid the effects from the previous pulse influencing the diffracted signal generated by the current pulse [Spence et al. 2012]. Given this minimum distance on the sample jet needed between shots, $d_{\min} = 50\mu$m, and the flow velocity of the jet, v_{jet}, the maximum repetition rate for the XFEL beam is then $v_{\text{jet}}/d_{\min}$. For a jet velocity of 10ms^{-1}, this gives an upper bound on the repetition rate of the XFEL pulse to be 200kHz (current repetition rate at the LCLS is 120Hz).

3.4.4 The X-ray Detector

With the unique properties of XFELs, a dedicated class of detectors are needed to properly capture results from the experiments. The pulsed nature and the high X-ray intensity of these sources requires the use of integrating rather than photon-counting detectors, furthermore detector frames resulting from individual pulses must be read out before the next pulse.

The detector set-up employed by the first SFX experiment at the LCLS has two sets of p-n junction charge-coupled device (pnCCD) detectors with one situated closer to the interaction region and one further downstream as shown in Fig. 3.11. Each set of detectors consists of two sub-panels of 512×1024 pixels, with each pixel $75\mu\text{m} \times 75\mu$m in size [Struder et al. 2010]. The front detector measures high-angle scattering and the back detector measures low-angle scattering to give high and low resolution data, respectively. The sub-panels are separated by a horizontal gap and diffraction patterns are read out after each pulse. Split panels are required not only to let the low-angle diffraction signal pass through to the back detector but also to allow the unscattered pulse to defocus and pass downstream to a beam dump. The pnCCD detector is currently still in use at the Atomic, Molecular and Optical science instrument (AMO) at the LCLS.

The current workhorse detector at the Coherent X-ray Imaging instrument (CXI) at the LCLS is the Cornell-SLAC Pixel Array Detector (CSPAD) [Philipp et al. 2010, Hart

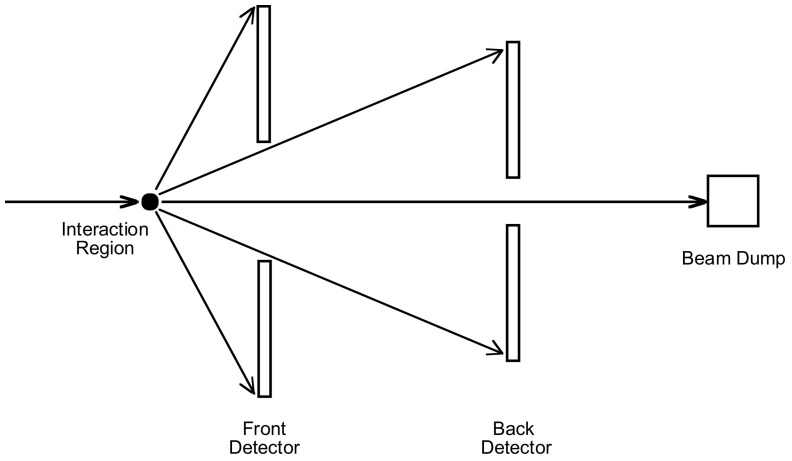

Figure 3.11 Layout of the detector set-up used in the first nanocrystallography experiment at LCLS.

et al. 2012, Carini et al. 2014]. The high density of electronic components in this detector necessitates the use of application-specific integrated circuits (ASICs). The detector consists of four quadrants with each quadrant being composed of eight 2×1 modules. Each 2×1 module is made of two ASICs, bump-bonded to a single 500μm thick silicon pixel array sensor where the centre of the module dividing the two ASICs is a dead-zone with a width of 3 pixels [Hart et al. 2012]. Each ASIC controls 185×193 pixels, with each pixel $110\mu\text{m} \times 110\mu\text{m}$ in size and forms the basic building block for the CSPAD. In total there are 64 ASICs on the CSPAD which makes it a 2.3Mpixel camera. The physical layout of the CSPAD is shown in Fig. 3.12(b) and (d).

The CSPAD quadrants are water-cooled and are also offset with respect to the centre of the detector to create a gap for the XFEL pulse to pass through. The quadrants can be moved radially in-situ to accommodate varying beam widths or to capture signals at different positions in reciprocal space.

Figure 3.12(a) and (c) shows the raw layout of the CSPAD when the data on the detector are read-out. Assembling a physically correct image requires interpolation of the raw data onto a regular pixel grid due to the moveable quadrants and mechanical tolerances in the placement of individual modules. Obtaining the physical configuration thus results in averaging of measurements, and for this reason, data analysis is performed in the raw layout whenever possible. An accurate detector geometry can be obtained in practice by calibrating with powder samples which generate circular diffraction rings as shown in Fig. 3.12(d) that can be used to align the different pieces of the detector.

Note that a previous version of the CSPAD had exhibited undesirable non-linearities that disrupted diffraction measurements for weak signals [Herrmann et al. 2013a]. The issue has recently been tracked down and found to be due to a combination of the high X-ray intensities and short charge collection times experienced by the detector [Carini et al. 2014]. A voltage build-up caused by these two factors lengthened the reset time

of components on the detector, resulting in charges leaking through. Thus the next version of this detector will need to synchronize the arrival of the diffracted signal with an additional delay.

Future detectors will need to deal with dynamic range issues and high pulse repetition rates of upcoming XFELs [Denes 2014]. Newer versions of the CSPAD with improved linearity and gain homogeneity have been constructed [Herrmann et al. 2013b] and the next generation detectors for experiments at the LCLS are also under development, called the "ePix" detector [Dragone et al. 2014]. Two-dimensional integrating X-ray detectors with read-out speeds of up to 4.5MHz are currently being developed to accommodate the 30kHz pulse repetition rate anticipated at the European XFEL [Koch et al. 2013].

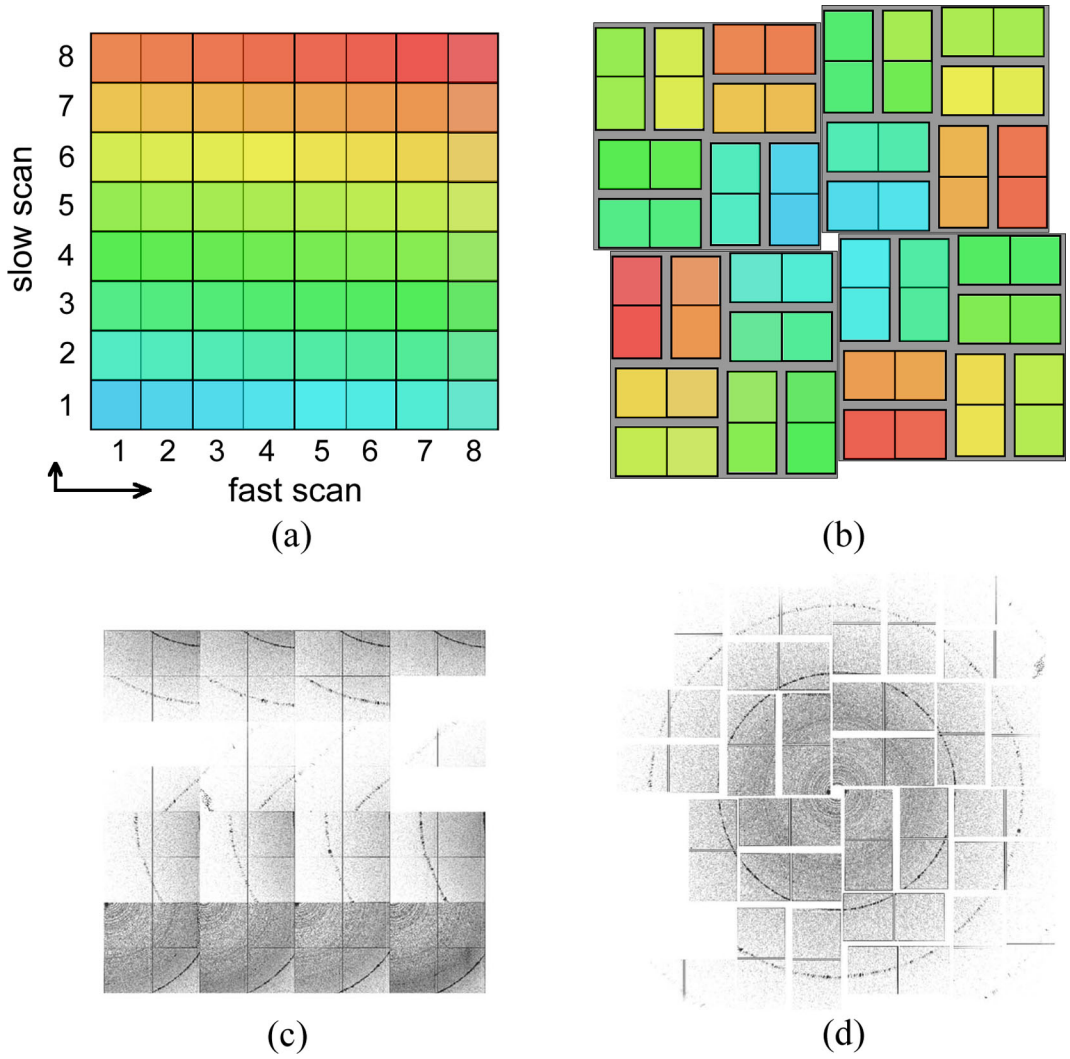

Figure 3.12 Layout of the CSPAD; adapted from Barty et al. [2014]. (a) Raw layout of the detector data as presented in a computer for data processing. (b) Physical layout of the CSPAD. The colours in (a) and (b) depict the order of detector read-out for the individual ASIC tiles. Blue tiles are the first to be read out and red tiles are the last. (c) and (d) corresponds to (a) and (b), respectively for an actual powder diffraction pattern.

3.4.5 Hit-finding, Data Transport and Storage

The 120Hz repetition rate of the XFEL pulse at the LCLS suggests that a typical SFX experiment that runs continuously for 5 days can give on the orders of 10^7 diffraction pattern frames per experiment. The European XFEL has the capacity to generate even more data than at the LCLS due to its higher repetition rate. Because the position, and also the orientation, of the crystals in the liquid delivery jet cannot be reliably controlled, not every detector read-out will contain diffraction from a single crystal. Many of the XFEL pulses will not intercept a crystal while others may meet two or more crystals during its traversal through the sample stream and will produce diffraction patterns from multiple and occluded crystals. Thus, efficiently reducing the obtained data in real-time is critical for both guiding users during experiments and for the rapid extraction of experimental outputs once the experiments have concluded. The so-called "hit-finding" software programs are the tools required for this data reduction to be done. Currently, all diffraction data frames at the LCLS are saved for off-line analysis and as mentioned above, only a small fraction of the measured data will contain usable, good quality diffraction in most cases. The hit-finding program analyses the stored data and retains only patterns that contain usable diffraction. The work-flow diagram of Fig. 3.13 depicts the potential that this data reduction step has to significantly reduce recorded data volumes.

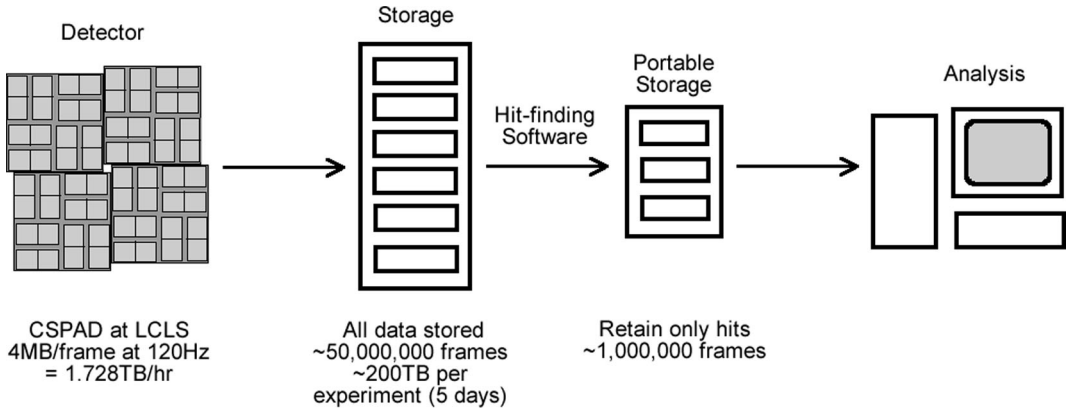

Figure 3.13 Flow diagram of the first stage of the data reduction process.

Specialised software have been developed to perform preliminary analysis and data reduction. One prominent program so-far is an open-source package called "Cheetah" [Barty et al. 2014] with its namesake derived from the need for speedy processing of the incoming data. The purpose of software such as Cheetah is to evaluate the quality of each data frame to provide real-time feedback on the experimental progress and to sort the diffraction data according to its quality. Data can be sorted according to various criteria and compiled into reduced forms such as powder patterns for subsequent analysis. The functions of Cheetah include:

- Correction for detector artefacts.
- Estimation and subtraction of background photons.
- Hit-finding and frame-sorting.
- Generation of statistics on hit-rate and resolution.
- Conversion of selected data into facility-independent formats for subsequent analysis.

Hit-finding and frame-sorting can be carried out by setting a threshold on the pixel intensities that is able to identify Bragg peaks and other known features of the particular diffraction pattern. For SFX experiments where nanocrystals are used, the number of Bragg peaks can then be used to gauge the quality of a diffraction pattern. Only frames suitable for the experiment at hand are retained, usually from single crystal hits for SFX. The frames without any hits can be summed to provide an estimate of the background noise and this estimate can be fed back into the program to subtract the frames with hits.

The current LCLS data retention policy is that all data are retained for 6 months, after which only a selection of the data are retained on disk. The typical delay between experiment and publication is roughly 2 years, with approximately 6 months for preliminary analysis, 6-8 months for a first paper draft, 4-8 months in review, and 2 months to publish [Barty et al. 2013]. At any point in this process it may be necessary to access the original data either to answer reviewer questions or to apply the latest analysis techniques for better results.

Hit-finding software such as Cheetah is facility-independent and can be readily adapted for use at different lasing facilities [Barty et al. 2014]. At the European XFEL, saving every single frame for post-processing will no longer be practical and data reduction will need to be performed before the frames are stored.

The reduction in data, and subsequently, the decrease in the transfer and storage burdens gives research groups who do not have access to large-scale computation or storage infrastructures an opportunity to still use XFELs for their investigations. Improvements to the pre-processing of data using programs such as Cheetah may allow old data to be re-analysed to obtain better results.

3.5 DIFFRACTION DATA ANALYSIS

The result from the data collection process is a data-set that must now be assembled and analysed to extract the structure of the molecule. The generic data analysis workflow for nanocrystals is to first map each detector pixel into their correct positions in reciprocal space and then to merge the intensity data via averaging to obtain a quantity

related to the Fourier transform of the unit cell (either the structure factor amplitudes, $|F_{\mathbf{h}}|$, or the continuous diffraction from one unit cell, $|F(\mathbf{u})|$). Inversion into real space can then be performed to recover the electron density of the unit cell. These steps are described in this section.

3.5.1 Orientation Determination

Since each diffraction pattern corresponds to a 2-D slice of the 3-D Fourier intensity of the crystal, many patterns from different crystal orientations are needed to build up the 3-D Fourier intensity of the crystal. The randomly oriented and large number of diffraction patterns resulting from SFX gives the many different crystal orientations needed. The basic principle to determine the orientations of the individual patterns is described below.

For a flat detector with its normal vector parallel to the incident beam and centred about the z-axis in which the undeflected incident beam propagates, the diffraction experiment can be depicted by Fig. 3.14. The unit scattering vectors $\mathbf{s}_i$ and $\mathbf{s}_o$, form the difference scattering vector $\mathbf{u}$ that indexes reciprocal space as described in Sec. 2.3.3. The two distances of interest are D_o, the sample-to-detector distance (also sometimes referred to as the *camera length*), and D_{ij}, the distance from the sample to the (i,j)th detector pixel. The position vector of the (i,j)th detector pixel with respect to the centre of the detector is denoted $\mathbf{p}_{ij}$. The goal is to express the reciprocal space vector $\mathbf{u}$ in terms of $\mathbf{p}_{ij}$.

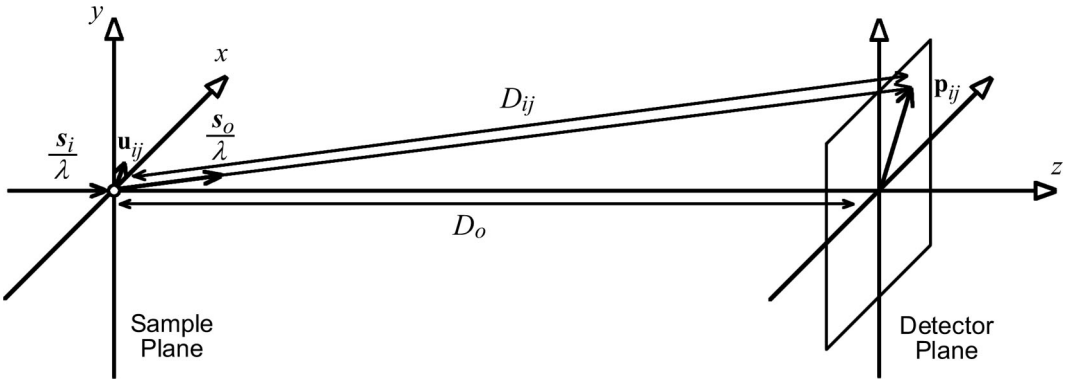

Figure 3.14 Geometry of a general diffraction experiment.

Recall from Sec. 2.3.3,

$$\mathbf{u} = \frac{1}{\lambda}\left(\mathbf{s}_o - \mathbf{s}_i\right). \tag{3.19}$$

Let $\hat{\mathbf{z}}$ be the unit vector in the z-direction, then from Fig. 3.14, the position of the (i,j)th pixel to the outgoing unit scattering vector is related by

$$\mathbf{s}_o = \frac{\mathbf{p}_{ij} + D_o\hat{\mathbf{z}}}{D_{ij}}, \tag{3.20}$$

and the incident unit scattering vector is simply

$$\mathbf{s}_i = \hat{\mathbf{z}}. \tag{3.21}$$

D_{ij} can be found through the relation $D_{ij} = \sqrt{D_o^2 + |\mathbf{p}_{ij}|^2}$ and the reciprocal space vector $\mathbf{u}_{ij}$ for the (i,j)th pixel can thus be written as

$$\mathbf{u}_{ij} = \frac{1}{\lambda D_{ij}}\big(\mathbf{p}_{ij} + (D_o - D_{ij})\,\hat{\mathbf{z}}\big). \tag{3.22}$$

In the case where the finite-sized detector is placed in the far-field, $D_{ij} \approx D_o$ for all i and j, resulting in the approximation

$$\mathbf{u}_{ij} \approx \frac{1}{\lambda D_o}\mathbf{p}_{ij}, \tag{3.23}$$

useful for determining the maximum resolution roughly obtainable from an experimental detector set-up.

Writing $\mathbf{u}_{ij} = \mathrm{L}^{-1}\mathbf{h}_{ij}$ as described in Sec. 2.3.5, where $\mathbf{h}_{ij}$ are the Miller indices corresponding to the reciprocal space position $\mathbf{u}_{ij}$ and substituting into Eq. (3.22), the intensities in a diffraction pattern may then be mapped into reciprocal space via

$$\mathbf{h}_{ij} = \frac{1}{\lambda D_{ij}}\mathrm{L}\big(\mathbf{p}_{ij} + (D_o - D_{ij})\,\hat{\mathbf{z}}\big), \tag{3.24}$$

which relates the detector pixel positions to the Miller indices. Usually the Miller indices resulting from this matrix inversion will not be integers, therefore depending on the details of the data integration process, the intensities can either be mapped into an up-sampled array or the Miller indices can be rounded.

The generator matrix L for the crystal lattice is typically determined through so-called *autoindexing algorithms* that extract the orientation information and the lattice constants of a crystal from the collected diffraction patterns. An efficient algorithm for finding the basis vectors of the lattice was developed by Rossmann and van Beek [1999] and utilises the Fourier transform to search for the periodicity of peak intensities projected onto a line in reciprocal space. The Fourier projection method was originally implemented in the crystallography software package *DPS* [Steller et al. 1997], now also used in popular program suites *MOSFLM* [Leslie 1991] and *d*TREK* [Pflugrath 1999]. *MOSFLM* originally employed an indexing method based on analysing the distribution of difference vectors between all Bragg peaks [Kabsch 1993]. Another popular program is *DENZO*, distributed as part of the *HKL* diffraction image processing suite [Otwinowski and Minor 1997]. For more information on the crystallographic autoindexing process and the integration of diffraction data in general, see Leslie [2006] and references for the various computer packages given above.

A major advantage of diffraction imaging using nanocrystals over that of single molecules is that the orientation problem is much more easily solved through crystallographic indexing. For nanocrystals with a very small number of unit cells, orientation determination methods for single particle may be used [Loh and Elser 2009, Fung et al. 2009, Bortel and Tegze 2011]. Note that these methods can also be used to determine the orientation of entire nanocrystals which may be utilised as a part of future autoindexing algorithms or at least form a useful independent check on existing indexing techniques. A general review on the orientation determination problem for nanocrystals and single particles in serial XFEL imaging can be found in Spence et al. [2012].

In summary, the 2-D diffraction patterns obtained from an SFX experiment are oriented in 3-D Fourier space by first finding and indexing the Bragg peaks in the patterns using well-established crystallographic indexing programs, called during the execution of dedicated data processing software for SFX. Two such SFX data analysis software developed so far are CrystFEL [White et al. 2012] and the Computational Crystallography Toolbox (Cctbx) [Sauter et al. 2013].

3.5.2 Intensity Merging

With the determination of the orientation for each diffraction pattern complete, the diffracted intensities from all patterns can be merged into a 3-D reciprocal space intensity map. The process typically involves creating a three-dimensional grid oversampled with respect to the Bragg peaks, binning the intensity values into their correct positions according to their orientations, and then averaging the intensities from each pattern if they fall within the same grid point [Kirian et al. 2010, Kirian et al. 2011, Yefanov et al. 2014]. Because only one diffraction pattern can be acquired from each crystal in an SFX experiment and all the crystals are distinct, this procedure of merging results in a summed diffracted intensity from crystals of different sizes and shapes. Averaging together many oriented nanocrystal diffraction data is given the name *Monte Carlo intensity merging* [Kirian et al. 2010] and has the desirable effect of averaging out experimental uncertainties in practice.

Monte Carlo intensity merging can be applied to yield two different but related quantities that can both be used to determine the electron density of the molecule. The first is the structure factor amplitudes $|F_{\mathbf{h}}|$ as in conventional X-ray crystallography, and the second is the continuous amplitude of the Fourier transform of the unit cell, $|F(\mathbf{u})|$. The method to obtain $|F_{\mathbf{h}}|$ is discussed below. The details and significance of the intensity merging process for obtaining $|F(\mathbf{u})|$ is given proper treatment in Chapters 4 and 5.

Let $I_n(\mathbf{u})$ denote the oriented intensity data from the nth diffraction pattern and $S_n(\mathbf{u})$ the corresponding shape transform. The determination of structure factor am-

plitudes $|F_{\mathbf{h}}|$ in an SFX experiment involves first defining an integration domain of radius δ in reciprocal space such that all intensities lying within a ball $|\mathbf{u}-\mathbf{u}_{\mathbf{h}}|<\delta$ are averaged together. For δ sufficiently small, the molecular transform can be assumed to be constant within the ball, and the diffracted intensity averaged in this way becomes

$$\left\langle I_n\left(\mathbf{u}-\mathbf{u}_{\mathbf{h}}\right)\right\rangle_{\delta,\mathbf{h},n} \approx |F_{\mathbf{h}}|^2\left\langle\left|S_n\left(\mathbf{u}-\mathbf{u}_{\mathbf{h}}\right)\right|^2\right\rangle_{\delta,\mathbf{h},n} \tag{3.25}$$

via Eq. (2.82), where $\langle\rangle_{\delta,\mathbf{h},n}$ indicates that the average is done over all patterns and only the intensity measurements satisfying $|\mathbf{u}-\mathbf{u}_{\mathbf{h}}|<\delta$ are included in the average. By Eq. (3.25), the structure factor amplitude is then

$$|F_{\mathbf{h}}|^2 \approx \frac{\left\langle I_n\left(\mathbf{u}-\mathbf{u}_{\mathbf{h}}\right)\right\rangle_{\delta,\mathbf{h},n}}{\left\langle\left|S_n\left(\mathbf{u}-\mathbf{u}_{\mathbf{h}}\right)\right|^2\right\rangle_{\delta,\mathbf{h},n}}. \tag{3.26}$$

Since the shape transform is the same around each reciprocal lattice point, $S_n(\mathbf{u})$ is invariant under translations $\mathbf{u}_{\mathbf{h}}$ for all $\mathbf{h}$ thus $\langle|S_n(\mathbf{u}-\mathbf{u}_{\mathbf{h}})|^2\rangle_{\delta,\mathbf{h},n} = \langle|S_n(\mathbf{u})|^2\rangle_{\delta,n}$ which also approaches a constant for δ sufficiently small. Therefore the structure factor amplitudes can be obtained to within a proportionality constant from the averaged diffracted intensity simply via

$$|F_{\mathbf{h}}|^2 \propto \left\langle I_n\left(\mathbf{u}-\mathbf{u}_{\mathbf{h}}\right)\right\rangle_{\delta,\mathbf{h},n}. \tag{3.27}$$

In practice the polarisation of the XFEL beam and the solid angle subtended by a detector pixel will also need to be taken into account. Furthermore, since all X-ray beams are neither monochromatic nor perfectly collimated, the Ewald sphere surface will therefore have a finite thickness with that thickness varying to give a crescent-like shape when the sphere is sliced in 2-D. A phenomenon called *partiality* can then arise where each Bragg reflection in a diffraction pattern intersects the Ewald shell only partially and each in a different way. The intensity of the Bragg reflections is therefore altered depending on the amount of their intersection with the imperfect Ewald sphere. Partiality can therefore be a serious problem when it comes to merging SFX data[1]. See White [2014] for further details.

Kirian [2011] noted that the value of δ should be chosen such that it is smaller than the finest features of the molecular transform by the sampling theorem, which corresponds to approximately $\delta < 1/(2d)$ where d is the maximum dimension of the unit cell. He also realised that the errors in the structure factor on a particular reciprocal lattice point is given by $\sigma(I_{\mathbf{h}})/\sqrt{N_{\mathbf{h}}}$ according to the central limit theorem where $\sigma(I_{\mathbf{h}})$ is the standard deviation of the intensity measurements and $N_{\mathbf{h}}$ is the total number of measurements averaged at a reciprocal lattice point $\mathbf{h}$. Thus δ too

[1]Partiality arises almost exclusively in crystallography with XFELs due to a single one and only one effectively instantaneous diffraction pattern being taken for each crystal, thus without the possibility of changing the experimental conditions that alleviates the problem in conventional X-ray crystallography.

small will result in a correspondingly small $N_{\mathbf{h}}$ as fewer intensity measurements would be found in the integration domain. The intensity merging process would then have a slower convergence, requiring more diffraction data. δ too large will increase the variance in the intensities by incorporating more background noise and empty pixels into the average. Moreover, a large δ would also break the assumption that the shape transform and the molecular transform are constant within the integration region. The breakage would come first for the shape transform as $|S_n(\mathbf{u})|^2$ varies more quickly than $|F(\mathbf{u})|^2$ in general. The verification of this intensity merging approach using actual SFX experimental data gave an optimum value of δ being approximately equal to the inverse of the mean size of the crystals [Kirian et al. 2010].

3.5.3 Inversion into Real Space - Phasing in SFX

The intensity merging process described above determines the structure factor amplitudes $|F_{\mathbf{h}}|$ which can be phased using molecular replacement as with conventional X-ray crystallography data. Phasing through MAD for crystallographic experiments using XFELs has also been investigated [Son et al. 2011, Son et al. 2013] and shown to be feasible provided time-dependent X-ray absorption coefficients could be obtained.

A method to phase SFX data using the continuous molecular transform amplitude $|F(\mathbf{u})|$ instead of the structure factor amplitudes have been proposed by Spence et al. [2011]. The basic procedure involves obtaining an estimate of $|F(\mathbf{u})|$ and using it to reconstruct the electron density of the unit cell via iterative phase retrieval algorithms. Kirian et al. [2015] have recently given the first experimental demonstration of this method and have shown that the approach works in practice and is relatively robust in terms of being insensitive to fluctuations in X-ray beam qualities. Understanding and applying this method is one of the main impetuses of this thesis and its analysis and justification are given in Chapter 5.

Elser [2013] has proposed an alternative but related approach that uses the intensity measurements as well as their gradients at the Bragg reflections to reconstruct the electron density from XFEL nanocrystal diffraction data. This approach avoids the use of the weak amplitudes between the Bragg reflections but still requires a noise-sensitive estimate of the intensity gradients. Much more will be said on utilising the continuous function $|F(\mathbf{u})|$ to perform phase retrieval in SFX in the coming chapters.

3.6 SFX METHOD SUMMARY

The work-flow of structure determination using SFX from data collection to inversion into real space as described in the previous few sections is summarised below:

1. Diffraction pattern collection.

2. Data reduction - Hit-finding, background subtraction.

3. Data analysis - Peak finding and orientation determination by indexing.

4. Structure factor determination - Intensity merging.

5. Phase retrieval.

6. Structural refinement and model-fitting.

At the completion of the structure determination process, diffraction data from collaborations associated with the LCLS are currently deposited in the Coherent X-ray Imaging Data Bank (CXIDB) [Maia 2012], a publicly accessible online repository. The CXIDB allows testing of new and existing algorithms to be carried out with real-world data that should help promote and drive the development of algorithms needed for imaging using XFELs.

3.7 CURRENT PROGRESS AND OUTSTANDING CHALLENGES

The first serial femtosecond nanocrystallography experiment was undertaken in 2009 at the LCLS [Chapman et al. 2011]. The target specimen was the protein group Photosystem I responsible for photosynthetic processes in plants and organisms such as algae. A succession of sub-micron nanocrystals of different sizes and shapes were streamed into the XFEL pulse through the GDVN delivery system and diffraction patterns of resolutions up to 8.5Å was recorded at a rate of 30 patterns per second. The duration of the X-ray pulses was varied in this experiment and the resulting data showed that pulse durations of 70fs or less were suitable to prevent radiation damage from being seen in the resulting data. This first proof-of-concept experiment was a great success in that damage-free diffraction patterns were shown to be obtainable using an XFEL and it also provided insight into the experiment and the data processing requirements for the SFX method.

A structure solved to atomic resolution was first achieved using SFX by Boutet et al. [2012]. Electron densities of the lysozyme protein from egg whites of both hens and turkeys were solved to 1.9Å resolution with the aim of the experiment to compare the well known structure of lysozyme with synchrotron data obtained using macroscopic crystals as a control. The assembled data from these experiments are shown to be consistent with the known structures of these molecules. Moreover, the structure of the lysozyme protein from hen and turkey was shown to be readily distinguishable, leading John Spence, one of the main pioneers of the SFX method, to quip in a meeting of the Royal Society in 2013 that these X-ray lasers are machines capable of telling different species of birds apart.

A previously unknown protein structure, *Trypanosoma brucei* cysteine protease cathepsin B, associated with sleeping sickness has been solved using SFX in conjunction with the molecular replacement method for phasing [Redecke et al. 2013]. Phase determination using heavy-atom techniques at an XFEL have also been shown to be feasible by Barends et al. [2013], demonstrating the possibility of recovering a structure that does not have a known homology model.

Structure determination from membrane proteins crystallised in the lipidic phases have now also been demonstrated [Johansson et al. 2012, Johansson et al. 2013], where it was shown that nanocrystals of membrane proteins could be grown in a LCP medium and injected directly into the XFEL pulse. Other membrane proteins determined by SFX to date include the Photosystem II complex to 5.7Å resolution [Kerna et al. 2012], and the human serotonin receptor, a type of the so-called G protein-coupled receptors, to 2.8Å resolution [Liu et al. 2013].

Building on the experiences garnered from the achievements described above, there are still much work that needs to be done. The sample delivery mechanisms of SFX are under intense research in order to optimise their performance while minimising jet instability and clogging. Detectors capable of fast data read-out rates with the ability to withstand the extreme conditions generated by XFEL pulses are also required as described in Sec. 3.4.4. Other challenges that need to be addressed in the general scheme of serial femtosecond X-ray imaging include:

- Dealing with the significant and highly variable background scattering that are present from the water envelope surrounding a hydrated sample.
- Development of efficient and parallelised algorithms that can sort through millions of patterns in real-time to separate out unusable data from valid diffractions.
- Having specialised equipment and techniques to handle the imaging of dangerous and volatile specimens such as contagions or harmful proteins.

Care also needs to be taken to ensure that systematic errors and biases such as a particular direction being favoured by the crystals during their delivery to the XFEL pulse does not occur or is taken into account. The bias could arise due to the force imparted on the samples from the directional flow of the liquid jet and such sample alignment would require the resulting data to be analysed differently.

Finally, each molecule to be imaged would ultimately be different, with those differences varying both in space and in time, which need to be carefully considered during the reconstruction process in order for any imaging technique to be able to obtain something beyond a spatially and temporally averaged answer.

Chapter 4

DIFFRACTION CHARACTERISTICS OF NANOCRYSTALS - COMPLETE UNIT CELLS

4.1 INTRODUCTION

The final obstacle of X-ray crystallography listed in the previous chapter was our current inability to accurately measure the phase of the diffracted X-ray wavefront, thus preventing a direct inversion of the diffraction pattern via the Fourier transform. This difficulty is exacerbated by the periodicity inherent in crystals, giving only diffraction signals concentrated at particular discrete locations called Bragg reflections. The measurable information in an X-ray crystallography experiment therefore consists of the magnitude of the molecular transform sampled at Bragg locations with no information between the Bragg samples. This undersampling of the Fourier transform amplitude prevents direct phase retrieval as described in Sec. 3.3.3.5.

The situation becomes slightly different for crystallography experiments carried out via the SFX method. The theory for some of the concepts needed to invert the diffraction intensities to retrieve an image of the unit cell is laid down in this chapter. Diffraction by small crystals is considered first and then the outcome of averaging together diffraction patterns from a collection of nanocrystals of different sizes is investigated. Only crystals consisting of an integral number of a single kind of complete unit cell is considered in this chapter. Effects due to different kinds of, and incomplete, unit cells that can occur when there is more than one molecule in the unit cell are discussed in Chapter 6.

4.2 DIFFRACTION BY A SINGLE CRYSTAL

Consider first a one-dimensional molecule residing inside a unit cell of length a on the interval $[-a/2, a/2)$ and let the electron density inside this unit cell be described by the function $f(x)$. A 1-D crystal of size N can be constructed by repeating the unit cell compactly, i.e., without gaps, N times along the axis and centred about the origin, as shown in Fig. 4.1. The electron density of the crystal, $g(N, x)$, can be expressed as

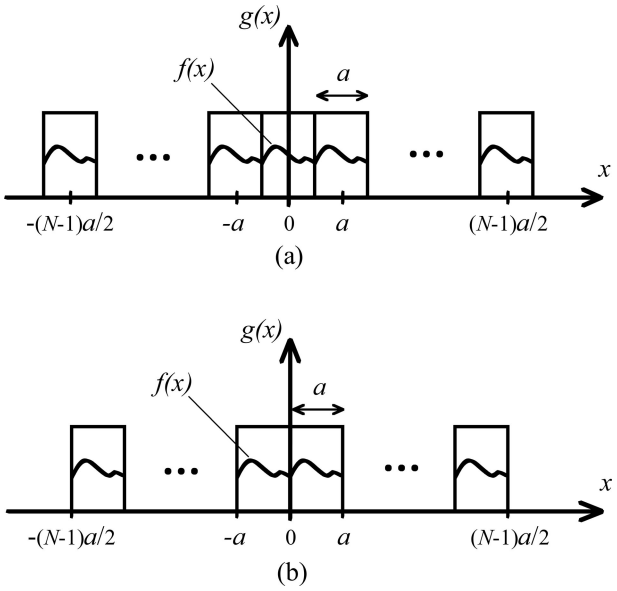

Figure 4.1 A one-dimensional crystal with N unit cells for (a) N odd, and (b) N even.

the convolution of $f(x)$ with a train of N delta functions spaced distance a units apart,

$$g(N,x) = f(x) \otimes \sum_{h=0}^{N-1} \delta\left(x - ha + \frac{1}{2}(N-1)\,a\right). \tag{4.1}$$

Note that since we have chosen the origin to be centred on the crystal, the position of the central unit cell(s) about the origin is different for N odd and for N even, as shown respectively in Fig. 4.1(a) and Fig. 4.1(b). The complex diffraction pattern in the far-field is proportional to $G(N,u)$, the Fourier transform of $g(N,x)$, so that by Eq. (4.1),

$$G(N,u) = F(u)\exp\left(i\pi(N-1)au\right)\sum_{h=0}^{N-1}\exp\left(-i2\pi hau\right), \tag{4.2}$$

with $F(u)$ being called here the *molecular transform*, i.e., the Fourier transform of one unit cell. Evaluating the geometric series in Eq. (4.2) simplifies the equation to

$$G(N,u) = F(u)S(N,u), \tag{4.3}$$

where

$$S(N,u) = \frac{\sin\left(N\pi au\right)}{\sin\left(\pi au\right)}. \tag{4.4}$$

The function $S(N,u)$ is referred to in this thesis as the *shape transform*. For N odd, $S(N,u)$ is known as the Dirichlet kernel (for N even, the peaks alternate in polarity). The intensity of the diffraction pattern from this crystal, $I(N,u)$, is given by

$$I(N,u) = |G(N,u)|^2 = |F(u)|^2\,S^2(N,u), \tag{4.5}$$

such that the crystal-size dependent function that modulates the intensity of the molecular transform is

$$S^2(N,u) = \frac{\sin^2(N\pi a u)}{\sin^2(\pi a u)}. \tag{4.6}$$

The function $S^2(N,u)$ is referred to here as the *squared shape transform*, in this case for a 1-D crystal of size N.

4.2.1 Dependence of $S(N,u)$ on N

The observed diffraction intensity for a crystal is therefore the squared magnitude of the molecular transform modulated by the squared shape transform. The effect of this modulation for different sized crystals is illustrated in Fig. 4.2. When $N = 1$, $S^2(N,u)$ is unity everywhere as expected to yield the diffraction from a single molecule $|F(u)|^2$, as shown in Fig. 4.2(a). In the limit as N tends to infinity, $S^2(N,u)$ converges to a comb function, modelling the situation in conventional crystallography using macroscopic crystals where only the Bragg reflections are observed, as illustrated in Fig. 4.2(c). In this situation, it is clear that for large crystals, the measurable information from a diffraction experiment is the intensity of the Fourier transform of one unit cell sampled at the Bragg positions. As the size of the crystal decreases, intensity between the Bragg peaks begin to appear albeit with the amplitude of the overall diffraction being much more attenuated. In general, there are $N-1$ zeroes between two adjacent Bragg positions as $S^2(N,u)$, and in turn, the diffracted intensity is zero for $u = k/(aN)$ where k is an integer.

Repeated application of L'Hospital's rule tells us that at the reciprocal lattice points where $u = u_h = h/a$, with h being an integer, the squared shape transform takes on the value

$$S^2(N,u_h) = N^2 \tag{4.7}$$

for all N even and odd. The diffracted intensity therefore scales as the squared number of unit cells at the reciprocal lattice points. At the half-integral reciprocal lattice points where $u = u_{h/2} = (2h-1)/(2a)$, we find

$$S^2(N,u_{h/2}) = \begin{cases} 1 & N \text{ odd} \\ 0 & N \text{ even.} \end{cases} \tag{4.8}$$

The diffracted intensity is therefore independent of N at half-integral reciprocal lattice points. Integrating over all u, the average diffracted intensity by Parseval's theorem is

$$\int_{-\infty}^{\infty} I(N,u)\mathrm{d}u = \int_{-\infty}^{\infty} |g(N,x)|^2 \mathrm{d}x, \tag{4.9}$$

and can be written subsequently using Eq. (4.1) as

$$\int_{-\infty}^{\infty} I(N,u)\mathrm{d}u = \int_{-\infty}^{\infty} \sum_{h_1=0}^{N-1} \sum_{h_2=0}^{N-1} f\left(x - h_1 a + \frac{1}{2}(N-1)\,a\right) f\left(x - h_2 a + \frac{1}{2}(N-1)\,a\right) \mathrm{d}x. \tag{4.10}$$

Since the unit cells are non-overlapping and $f(x-h_1a)f(x-h_2a) = 0$ for all $h_1 \neq h_2$, the integrated intensity is then

$$\begin{aligned}\int_{-\infty}^{\infty} I(N,u)\mathrm{d}u &= \int_{-\infty}^{\infty} \sum_{h=0}^{N-1} f^2\left(x - ha + \frac{1}{2}(N-1)\,a\right)\mathrm{d}x \\ &= \sum_{h=0}^{N-1} \int_{-\infty}^{\infty} f^2\left(x - ha + \frac{1}{2}(N-1)\,a\right)\mathrm{d}x \\ &= N\int_{-\infty}^{\infty} f^2(x)\,\mathrm{d}x. \end{aligned} \tag{4.11}$$

The diffracted intensity therefore scales as N when averaged over all reciprocal space. In summary, the diffracted intensity scales as N^2 at the reciprocal lattice points, N when averaged over all reciprocal space positions, and is independent of N at half-integral reciprocal lattice points.

4.2.2 Alternative Formulation using sinc Functions

An alternative expression to Eq. (4.3) for the diffraction by a finite crystal can be obtained by considering the finite crystal as a truncated segment of an infinite crystal. Modelling the truncation process as a multiplication by a rectangle function, the electron density of a 1-D crystal of size N unit cells can be written as

$$g(N,x) = \begin{cases} g_\infty(x)\ \mathrm{rect}\,(x/Na) & N \text{ odd} \\ g_\infty(x - a/2)\ \mathrm{rect}\,(x/Na) & N \text{ even,} \end{cases} \tag{4.12}$$

where $g_\infty(x)$ denotes the electron density of an infinite crystal with a unit cell centred about the origin as in Fig. 4.1(a). Taking the Fourier transform of Eq. (4.12), the diffraction can be expressed as the convolution of the Bragg reflections from an infinite crystal with a sinc function. The complex diffracted amplitude for a finite crystal of size N is therefore

$$G(N,u) = \begin{cases} G_\infty(u) \otimes \big(Na\ \mathrm{sinc}\,(Nau)\big) & N \text{ odd} \\ \big(G_\infty(u)\exp(-i\pi au)\big) \otimes \big(Na\ \mathrm{sinc}\,(Nau)\big) & N \text{ even,} \end{cases} \tag{4.13}$$

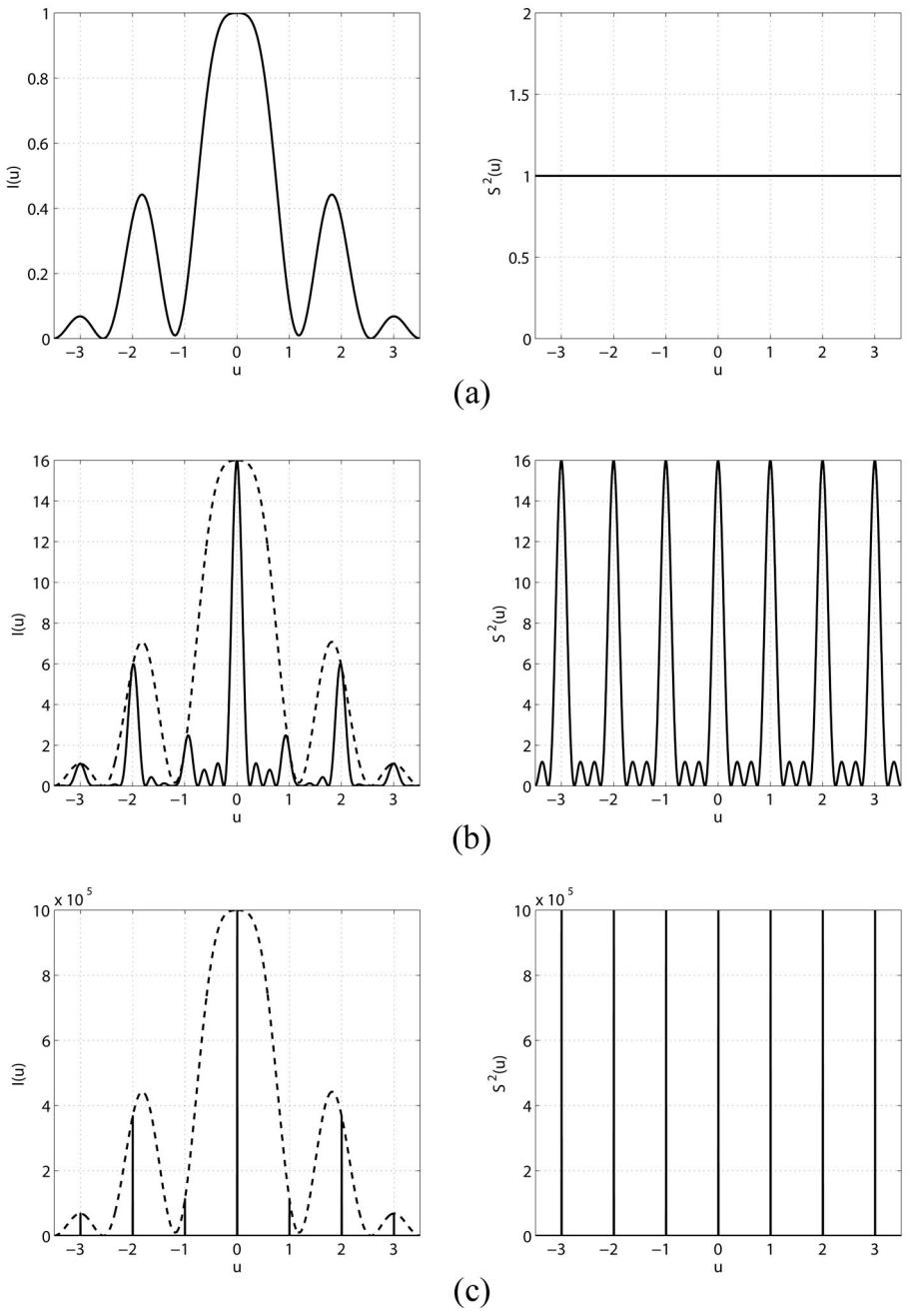

Figure 4.2 Diffracted intensity (left column) and its associated squared shape transform (right column) for a one-dimensional crystal of (a) 1 unit cell, (b) 4 unit cells, and (c) 1000 unit cells. The envelope (dashed line) in (b) and (c) is the squared amplitude of the molecular transform. The reciprocal space axis is normalised such that the Bragg peaks lie on integer values.

where $G_\infty(u)$ is the complex diffracted amplitude for an infinite crystal. Since $G_\infty(u)$ is the Bragg diffraction pattern, we have that

$$G_\infty(u) = \frac{1}{a}F(u)\sum_{h=-\infty}^{\infty}\delta\left(u-\frac{h}{a}\right). \tag{4.14}$$

Substituting Eq. (4.14) into Eq. (4.13) and combining the N odd and even cases gives

$$G(N,u) = N\sum_{h=-\infty}^{\infty}(-1)^{h(N-1)}F\left(\frac{h}{a}\right)\operatorname{sinc}\left(Na\left(u-\frac{h}{a}\right)\right). \tag{4.15}$$

Equation (4.15) is an alternative expression to Eq. (4.3) for $G(N,u)$, and shows that the diffraction profile around each Bragg reflection is a weighted sum of sinc functions.

The equivalence between Eqs. (4.3) and (4.15) is not obvious, but since their left-hand-sides are the same, the following mathematical identity must be true for all $F(u)$:

$$N\sum_{h=-\infty}^{\infty}(-1)^{h(N-1)}F\left(\frac{h}{a}\right)\operatorname{sinc}\left(Na\left(u-\frac{h}{a}\right)\right) \equiv F(u)\frac{\sin(N\pi au)}{\sin(\pi au)}. \tag{4.16}$$

Equation (4.16) can be shown directly as follows. Rearranging Eq. (4.16) and defining a quantity $X(u)$ as

$$X(u) = \frac{\sin(\pi au)}{\sin(N\pi au)}\,N\sum_{h=-\infty}^{\infty}(-1)^{h(N-1)}F\left(\frac{h}{a}\right)\operatorname{sinc}\left(Na\left(u-\frac{h}{a}\right)\right), \tag{4.17}$$

and keeping in mind the identity $\cos(\pi k) = (-1)^k$ for all $k \in \mathbb{Z}$, we have

$$\begin{aligned}
X(u) &= \frac{\sin(\pi au)}{\sin(\pi Nau)}\sum_{h=-\infty}^{\infty}\cos(\pi Nh-\pi h)\,F\left(\frac{h}{a}\right)\frac{\sin(\pi Nau-\pi Nh)}{\pi a\left(u-\frac{h}{a}\right)}\\
&= \frac{\sin(\pi au)}{\sin(\pi Nau)}\sum_{h=-\infty}^{\infty}\big(\cos(\pi Nh)\cos(\pi h)+\sin(\pi Nh)\sin(\pi h)\big)\,F\left(\frac{h}{a}\right)\\
&\qquad\frac{\sin(\pi Nau)\cos(\pi Nh)-\cos(\pi Nau)\sin(\pi Nh)}{\pi a\left(u-\frac{h}{a}\right)}\\
&= \frac{\sin(\pi au)}{\sin(\pi Nau)}\sum_{h=-\infty}^{\infty}\big(\cos(\pi Nh)\cos(\pi h)\big)\,F\left(\frac{h}{a}\right)\frac{\sin(\pi Nau)\cos(\pi Nh)}{\pi a\left(u-\frac{h}{a}\right)}\\
&= \sum_{h=-\infty}^{\infty}F\left(\frac{h}{a}\right)\frac{\sin(\pi au)\cos(\pi h)}{\pi a\left(u-\frac{h}{a}\right)}\\
&= \frac{1}{2}\sum_{h=-\infty}^{\infty}F\left(\frac{h}{a}\right)\frac{\sin(\pi a(u+\frac{h}{a}))+\sin(\pi a(u-\frac{h}{a}))}{\pi a\left(u-\frac{h}{a}\right)}\\
&= \frac{1}{2}\left(\sum_{h=-\infty}^{\infty}F\left(\frac{h}{a}\right)\frac{\sin(\pi a(u+\frac{h}{a}))}{\pi a\left(u-\frac{h}{a}\right)}+\sum_{h=-\infty}^{\infty}F\left(\frac{h}{a}\right)\frac{\sin(\pi a(u-\frac{h}{a}))}{\pi a\left(u-\frac{h}{a}\right)}\right).
\end{aligned} \tag{4.18}$$

Using the sampling relation of Shannon,

$$F(u) = \sum_{h=-\infty}^{\infty} F\left(\frac{h}{a}\right) \frac{\sin(\pi a(u - \frac{h}{a}))}{\pi a\left(u - \frac{h}{a}\right)}, \tag{4.19}$$

and noting that

$$\sin(\theta + \pi k) = \sin(\theta - \pi k) \tag{4.20}$$

for all $k \in \mathbb{Z}$ and any θ, Eq. (4.18) reduces to

$$X(u) = \frac{1}{2}\left(F(u) + F(u)\right) = F(u) \tag{4.21}$$

Putting this result into Eq. (4.17) and rearranging proves Eq. (4.16) as required.

The particular case when $F(u) = 1$ leads to the following identity:

$$N \sum_{h=-\infty}^{\infty} (-1)^{h(N-1)} \mathrm{sinc}\left(Na\left(u - \frac{h}{a}\right)\right) = \frac{\sin(\pi N a u)}{\sin(\pi a u)}. \tag{4.22}$$

This identity can be shown alternatively as follows. Denoting the left-hand-side of Eq. (4.22) by $L(u)$, substituting for the sinc function and considering the cases for N odd and even shows that

$$L(u) = -\frac{1}{\pi}\sin(\pi N a u)\left(\frac{1}{au} + 2au\sum_{h=0}^{\infty}(-1)^h \frac{1}{h^2 - (au)^2}\right). \tag{4.23}$$

The summation can be evaluated using the identity given by Eq. (4.103) of Wheelon [1968]:

$$\sum_{h=0}^{\infty}(-1)^h \frac{1}{h^2 - a^2} = -\frac{\pi}{2a}\left(\mathrm{cosec}\,(\pi a) + \frac{1}{\pi a}\right), \tag{4.24}$$

which shows that

$$L(u) = \frac{\sin(\pi N a u)}{\sin(\pi a u)}, \tag{4.25}$$

verifying Eq. (4.22) as required.

Although Eq. (4.15) implies the reflection profiles are a sum of weighted sinc functions, if the crystallites are sufficiently large, then there is minimal overlap between them, and the profile in the vicinity of each Bragg reflection reduces simply to $\mathrm{sinc}\,(Nau)$. The effect of the overlap can be approximately evaluated by considering the relative contribution from just the two neighbouring peaks, denoted here by ϵ. At the reciprocal lattice point $u = h/a$,

$$\epsilon = \frac{1}{F\,(h/a)}\left(F\left(\tfrac{h+1}{a}\right)\mathrm{sinc}\left(Na\left(\tfrac{h}{a} - \tfrac{h+1}{a}\right)\right) + F\left(\tfrac{h-1}{a}\right)\mathrm{sinc}\left(Na\left(\tfrac{h}{a} - \tfrac{h-1}{a}\right)\right)\right), \tag{4.26}$$

giving

$$\epsilon = \text{sinc}\,(N)\,\frac{1}{F\,(h/a)}\left(F\left(\tfrac{h+1}{a}\right) + F\left(\tfrac{h-1}{a}\right)\right). \tag{4.27}$$

The relative contribution from neighbouring reflections depends on their amplitudes relative to $F(h/a)$. However, if we assume that $F\,((h+1)/a) \approx F\,((h-1)/a) \approx F\,(h/a)$, then

$$\epsilon \approx 2\,\text{sinc}\,(N)\,. \tag{4.28}$$

The maximum error results when $\sin(\pi N) = \pm 1$ so that

$$\epsilon_{\max} \approx \frac{2}{\pi N}. \tag{4.29}$$

A relative contribution of less than 5% of the value of the structure factor at $u = h/a$ requires N to be approximately more than 13 unit cells. In three-dimensions the contributions come from six neighbouring structure factors, resulting in $\epsilon_{\max} \approx 6/\pi N$ and N greater than about 40 unit cells in each direction would be required to achieve less than 5% relative error. The contribution will be larger if the neighbouring reflections are stronger. This analysis shows that for nanocrystals on the order of 10 unit cells, it is necessary to use the full shape function $S(N, u)$ to describe the diffraction, rather than just a single sinc function.

4.3 ENSEMBLE-AVERAGED DIFFRACTION

In femtosecond nanocrystallography, diffraction patterns are collected from many crystals of varying sizes as they pass randomly across the pulsing XFEL beam. Assuming that the orientation of each pattern can be determined through automated Miller indexing as described in the previous chapter, the 2-D diffraction patterns can be classified, averaged, and assembled into a 3-D data volume in Fourier space. The variable instigating the most change in each diffraction pattern is the size of the crystal, i.e. the number of unit cells. Thus we now consider a description of the averaged diffraction from an ensemble of crystals.

4.3.1 The Averaged Shape Transform

The averaged diffracted intensity resulting from the data merging process can be described as follows. Let $I(N_n, u)$, $S(N_n, u)$ and N_n be the diffracted intensity, shape transform and the number of unit cells, for the nth nanocrystal, respectively. Writing $I(N_n, u)$ as $I_n(u)$ and $S(N_n, u)$ as $S_n(u)$, the diffracted intensity from the nth crystal is

$$I_n\,(u) = |F\,(u)\,|^2 S_n^2\,(u)\,. \tag{4.30}$$

The averaging process can now be written as

$$\left\langle I_n\left(u\right)\right\rangle_n = \left\langle \left|F\left(u\right)\right|^2 S_n^2\left(u\right)\right\rangle_n. \tag{4.31}$$

Since the molecular transform is identical for all snapshot diffraction patterns, we have

$$\left\langle I_n\left(u\right)\right\rangle_n = \left|F\left(u\right)\right|^2 \left\langle S_n^2\left(u\right)\right\rangle_n. \tag{4.32}$$

Over a large number of diffraction patterns from crystals of different sizes, the averaged diffracted intensity can also be written as

$$\left\langle I_n(u)\right\rangle_n = \sum_N p(N) I(N,u), \tag{4.33}$$

where $p(N)$ is the probability density function for the crystal size N. Substituting for the diffracted intensity of a single crystal in Eq. (4.33) using Eq. (4.5) allows us to write

$$\left\langle I_n(u)\right\rangle_n = \left|F(u)\right|^2 Q^2(u), \tag{4.34}$$

where

$$Q^2(u) = \left\langle S_n^2(u)\right\rangle_n = \sum_N p(N) S^2(N,u) = \sum_N p(N) \frac{\sin^2\left(N\pi a u\right)}{\sin^2\left(\pi a u\right)}, \tag{4.35}$$

which is referred to in this thesis as the *averaged squared shape transform*.

Figure 4.2(b) shows a plot of $\langle I_n(u)\rangle_n$ with the sizes of the nanocrystals following a truncated Gaussian distribution with a mean of 5 unit cells and a standard deviation of 5/3 unit cells. The truncation is due to the requirement that N must be greater than or equal to one, and in this case σ is chosen to be small enough such that the size distribution is essentially Gaussian. Note that by summing over the shape transform for various-sized crystals, the zeroes of $S^2(N,u)$ have been averaged out, making $Q^2(u)$ strictly positive for all u.

The maxima of the averaged squared shape transform occur at the reciprocal lattice points. Evaluating $Q^2\left(u\right)$ via Eq. (4.35) at $u = u_h$, and recalling that $S^2(N,u_h) = N^2$, yields

$$Q^2\left(u_h\right) = \sum_N p\left(N\right) N^2, \tag{4.36}$$

i.e., $Q^2\left(u\right)$ at the Bragg peaks is equal to the second raw moment of $p(N)$. For a Gaussian distribution of crystallite sizes, taking the continuum limit for N, we therefore have

$$Q^2\left(u_h\right) \approx \mu^2 + \sigma^2, \tag{4.37}$$

where μ and σ are the mean and standard deviation of the crystallite size distribution,

respectively. For example, for the case above, $\sigma = \mu/3$, which results in $Q^2(u_h) \approx 1.1\mu^2$. In general, unless the crystallite size distribution is quite broad, $Q^2(u_h) \approx \mu^2$.

The minimum values of the averaged shape transform occur halfway between adjacent Bragg peaks at $u_{h/2} = (2h-1)/(2a)$. Substituting this value of u into Eq. (4.35), gives

$$Q^2\left(u_{h/2}\right) = \sum_{N \text{ odd}} p\left(N\right). \tag{4.38}$$

It is reasonable to assume that $\sum_{N \text{ odd}} p(N) \approx \sum_{N \text{ even}} p(N)$ and since $\sum_N p(N) = 1$, the minimum value of the averaged shape transform is

$$Q^2\left(u_{h/2}\right) \approx 1/2 \tag{4.39}$$

for any $p\left(N\right)$.

4.3.2 The Inverse Averaged Shape Transform

The multiplicative inverse of the averaged squared shape transform is of relevance when estimating the continuous molecular transform from the averaged diffracted intensity. Using Eq. (4.35) shows that

$$\frac{1}{Q^2\left(u\right)} = 2\sin^2\left(\pi a u\right)\left(1 - \sum_N p\left(N\right)\cos\left(2\pi N a u\right)\right)^{-1}. \tag{4.40}$$

The summation in Eq. (4.40) approximates a Poisson sum and therefore tends to vanish except when the distribution $p(N)$ is so narrow such that there are an insufficient number of cosine terms in the summation for cancellation. Therefore, we expect that except for narrow crystallite size distributions,

$$\frac{1}{Q^2\left(u\right)} \approx 2\sin^2\left(\pi a u\right). \tag{4.41}$$

The function $1/Q^2(u)$ and its approximation given by Eq. (4.41) are calculated and shown in Fig. 4.3(b) for two Gaussian crystal size distributions of the same mean but with different standard deviations. It can be seen that the inverse averaged shape transform is closely approximated by Eq. (4.41), except for deviations occurring for the narrower size distribution.

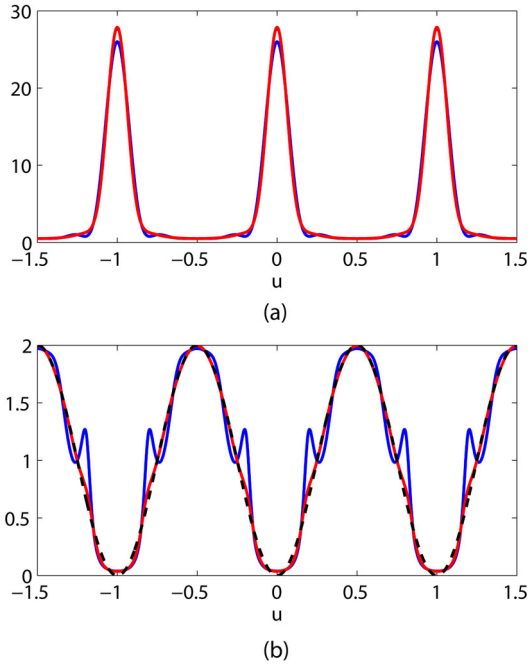

Figure 4.3 (a) The averaged squared shape transform and (b) its inverse, calculated for two Gaussian size distributions both of mean 5 unit cells but one with standard deviation 5/3 unit cells (red) and the other 1 unit cell (blue). The sinusoidal approximation for the inverse averaged shape transform given by Eq. (4.41) is shown by the dashed line in (b).

4.3.3 Effect of Crystal Size Distributions

The discussion so far has restricted the probability distribution for the crystallite sizes $p(N)$ to truncated Gaussians only. Here we investigate the effect that other distributions have on the averaged shape transform.

Choosing $p(N)$ to be uniform, truncated Gaussian and log-normal, the averaged squared shape transform for these three different distributions is shown in Fig. 4.4. The mean and standard deviation of each distribution are set to the same values for all three types of distributions to enable a fair comparison. The log-normal distribution is likely to be the most realistic distribution out of the three for describing real crystal sizes, as this distribution has been successfully applied to model the size distribution of aerosols and other small particles that result from the aggregation of small units building blocks [Limpert et al. 2001]. However, according to Fig. 4.4, the resulting shape of $Q^2(u)$ for both the uniform and log-normal distribution is found to be not much different than from using a truncated Gaussian. Thus, the overall behaviour of the averaged shape transform function seems to be little influenced by the particular crystal size distribution imposed and given that the truncated Gaussian is easy to calculate, it will continue to be used throughout the next chapter when conducting simulations using the averaged shape transform.

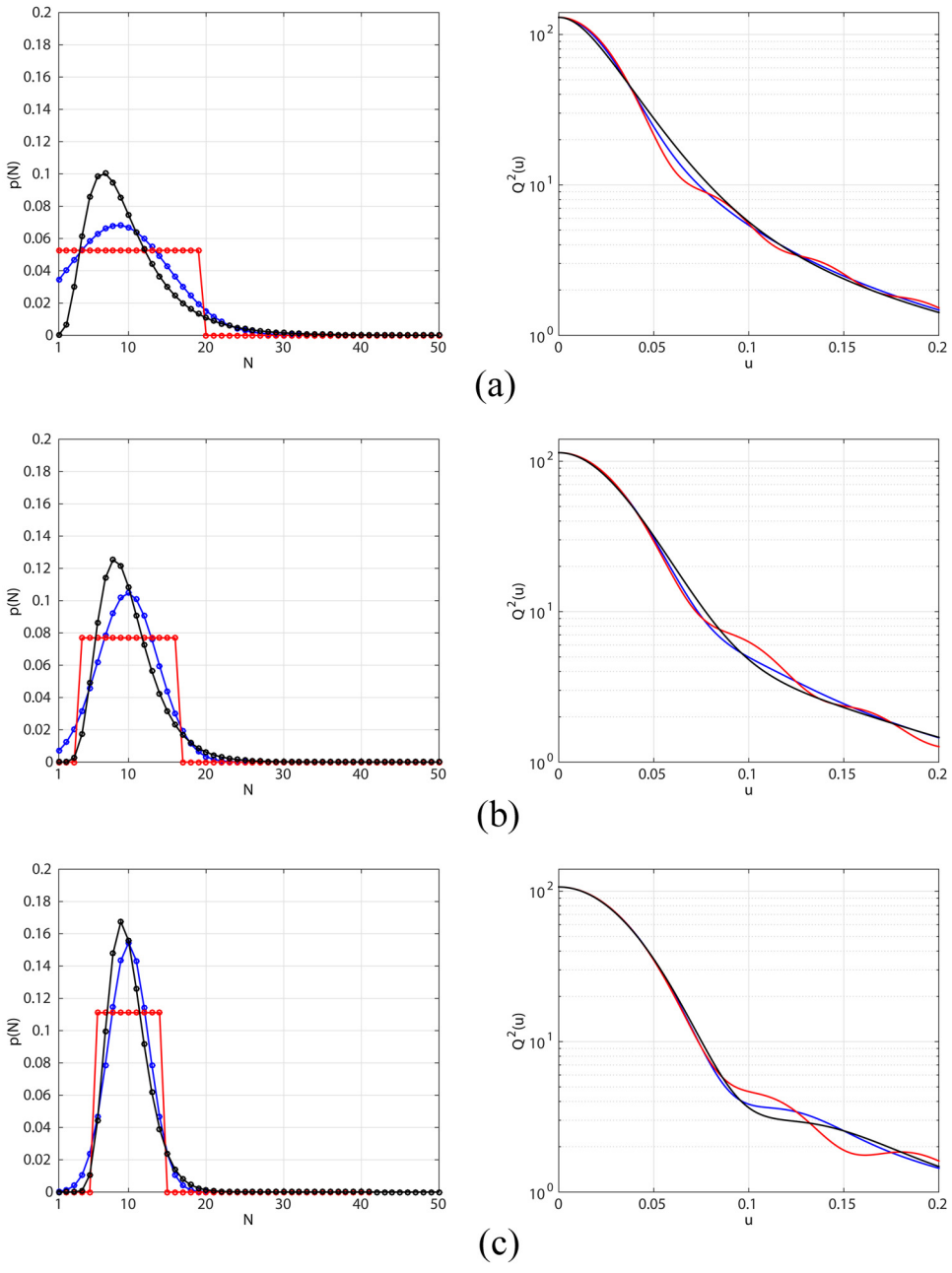

Figure 4.4 The averaged squared shape transform for three different crystal size distributions - uniform (red), truncated Gaussian (blue), log-normal (black), all with a mean of 10 unit cells and a standard deviation of (a) 5.48, (b) 3.74 and (c) 2.58 unit cells.

4.4 THE THREE-DIMENSIONAL CASE

In reality the crystals are three-dimensional. Following the derivations above, for a crystal of $N_1 \times N_2 \times N_3$ unit cells, the diffracted complex amplitude is given by

$$G(\mathbf{N}, \mathbf{u}) = F(\mathbf{u})S(\mathbf{N}, \mathbf{u}), \tag{4.42}$$

where $\mathbf{N} = [N_1, N_2, N_3]^T$ is the column vector of the number of unit cells in each of the crystal lattice directions given by the basis vectors $\mathbf{a}_1$, $\mathbf{a}_2$ and $\mathbf{a}_3$. The reciprocal space position vector that indexes Fourier space in the three reciprocal directions, can be decomposed into $\mathbf{u} = [u_1, u_2, u_3]^T$. $S(\mathbf{N}, \mathbf{u})$ is now the three-dimensional shape transform, given by

$$S(\mathbf{N}, \mathbf{u}) = \prod_{i=1}^{3} \frac{\sin(\pi N_i \mathbf{a}_i \cdot \mathbf{u})}{\sin(\pi \mathbf{a}_i \cdot \mathbf{u})}, \tag{4.43}$$

and the diffracted intensity from one single 3-D crystal is then

$$I(\mathbf{N}, \mathbf{u}) = |F(\mathbf{u})|^2 S^2(\mathbf{N}, \mathbf{u}). \tag{4.44}$$

Consider now a distribution of crystal sizes with PDF $p(\mathbf{N}) = p(N_1, N_2, N_3)$ such that $p(\mathbf{N})\mathrm{d}\mathbf{N}$ is the probability of a crystal size falling within the interval $(N_1, N_1 + \mathrm{d}N_1)$, $(N_2, N_2 + \mathrm{d}N_2)$ and $(N_3, N_3 + \mathrm{d}N_3)$. A wide range of forms for $p(\mathbf{N})$ is theoretically possible but we assume here for simplicity, that the PDF is jointly normal, i.e.,

$$p(\mathbf{N}) = \frac{1}{\sqrt{(2\pi)^3 \det(\mathrm{C})}} \exp\left(-\frac{1}{2}(\mathbf{N} - \boldsymbol{\mu})^T \mathrm{C}^{-1}(\mathbf{N} - \boldsymbol{\mu})\right) \tag{4.45}$$

where $\boldsymbol{\mu} = [\mu_1, \mu_2, \mu_3]^T$ is a column vector of the mean crystal sizes in each crystal axis direction, and C is the 3-by-3 covariance matrix given by

$$\mathrm{C} = \begin{pmatrix} \sigma_1^2 & \sigma_{12}^2 & \sigma_{13}^2 \\ \sigma_{12}^2 & \sigma_2^2 & \sigma_{23}^2 \\ \sigma_{13}^2 & \sigma_{23}^2 & \sigma_3^2 \end{pmatrix}. \tag{4.46}$$

If the crystal lengths N_1, N_2 and N_3 are uncorrelated with each other, i.e. the correlation coefficients $\rho_{ij} = \sigma_{ij}^2/\sigma_i\sigma_j$ satisfy $\rho_{12} = \rho_{13} = \rho_{23} = 0$, then the covariance matrix is diagonal. Because a joint Gaussian distribution is assumed, the length of a crystal in one direction when the correlation coefficients are zero is independent of the length in the other directions. In this case, the PDF factorises as

$$p(\mathbf{N}) = p(N_1)\, p(N_2)\, p(N_3), \tag{4.47}$$

where

$$p\left(N_i\right) = \frac{1}{\sqrt{2\pi}\sigma_i} \exp\left(-\frac{\left(N_i - \mu_i\right)^2}{2\sigma_i^2}\right), \tag{4.48}$$

for $i = 1, 2, 3$.

In general, we may suppose that complete independence of the crystal lengths is relatively unlikely since a crystal that grows in one direction is likely to grow in the other directions as well, i.e. needle-like or high aspect ratio crystals are special. Therefore, it is more probable that the crystal sizes N_1, N_2 and N_3 are positively correlated. At the other extreme, consider the case where the crystal sizes in each direction are completely correlated, i.e. the correlation coefficients satisfy $\rho_{12} = \rho_{13} = \rho_{23} = 1$. The joint density is then degenerate with only one degree of freedom, i.e. an instance of one random variable fixes the outcome of the other two. Without loss of generality, assume that N_1 is known, drawn from the single-variable Gaussian distribution given by Eq. (4.48) for $i = 1$, then the numbers of unit cells in the other two directions are given by

$$N_i = \frac{\sigma_i}{\sigma_1}\left(N_1 - \mu_1\right) + \mu_i, \tag{4.49}$$

for $i = 2, 3$. Equation (4.49) can be written as $(N_i - \mu_i)/(N_1 - \mu_1) = \sigma_i/\sigma_1$ which shows that the ratio of the difference between the length of each crystal edge and its mean is constant. However, it is also unlikely that the length of one crystal edge will fix the other two edges so it is reasonable that the correlation coefficients satisfy $0 < \rho_{ij} < 1$ in practice. Regardless of the detailed form of $p(\mathbf{N})$, for a collection of crystals the averaged squared shape transform in 3-D is given by

$$Q^2\left(\mathbf{u}\right) = \sum_{\mathbf{N}} p\left(\mathbf{N}\right) S^2\left(\mathbf{N}, \mathbf{u}\right). \tag{4.50}$$

Example crystallite size PDFs for a two-dimensional crystal and the resulting averaged shape transforms are shown in Fig. 4.5 for $\mu_1 = \mu_2 = 10$, $\sigma_1 = \sigma_2 = 3$, and for three values of the correlation coefficient. Figure 4.5 shows that increasing the correlation between the number of unit cells along each side of the crystal has only a small effect on the averaged shape transform. This is shown more clearly in Fig. 4.6 which shows a cut along $u_2 = 1$ through the centre of the shape transform. Increasing the correlation slightly sharpens the interference function, but the effect is small.

The assumption that the number of unit cells on each side of the crystal are independent means that $p(\mathbf{N})$ is of the form given by Eq. (4.47), and the averaged shape transform given in Eq. (4.50) is separable and can be written as the product of three one-dimensional averaged squared shape transforms, i.e.

$$Q^2\left(\mathbf{u}\right) = \prod_{i=1}^{3} Q^2(u_i) = \prod_{i=1}^{3} \sum_{N_i} p\left(N_i\right) S^2\left(N_i, u_i\right). \tag{4.51}$$

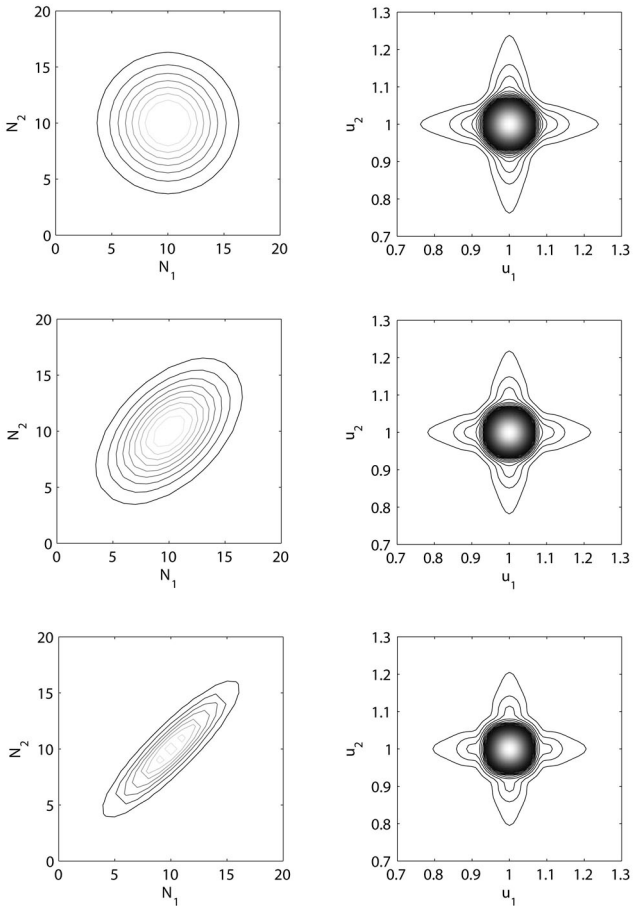

Figure 4.5 Contour plots of the PDFs (left column) and corresponding averaged squared shape transforms $Q^2(u_1, u_2)$ (right column) for a crystal size distribution with $\rho_{12} = 0$ (top row), $\rho_{12} = 0.5$ (middle row), $\rho_{12} = 0.9$ (bottom row).

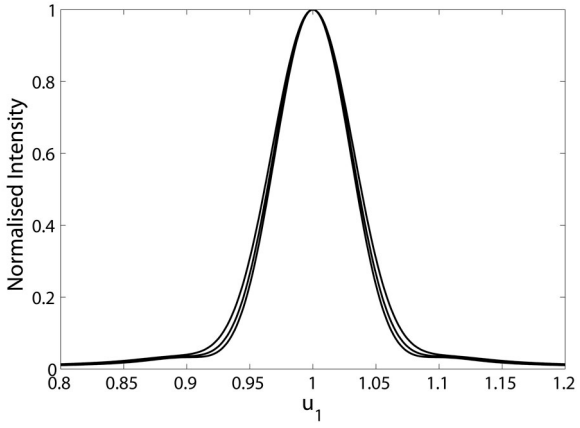

Figure 4.6 Normalised averaged shape transform profiles $Q^2(u_1, 1)$ for $\rho = 0$ (top curve), $\rho = 0.5$ (centre curve) and $\rho = 0.9$ (bottom curve).

The results derived in Sec. 4.3.2 for the maxima and minima of 1-D averaged shape transforms can easily be extended to three-dimensions with the independent side-lengths assumption as follows. Let $\mathbf{u_h}$ and $\mathbf{u_{h/2}}$ be the Bragg and half-integer Bragg positions in three-dimensions, respectively, as defined in Eqs. (2.86) and (3.14), the maxima of $Q^2(\mathbf{u})$ then occur when $\mathbf{u} = \mathbf{u_h}$ and have the value

$$Q^2\left(\mathbf{u_h}\right) = \sum_{\mathbf{N}} p\left(\mathbf{N}\right) N_1^2\, N_2^2\, N_3^2, \tag{4.52}$$

which is approximately equal to $(\mu_1\, \mu_2\, \mu_3)^2$ for a Gaussian distribution that is not too broad. The minima of $Q^2(\mathbf{u})$ occur when $\mathbf{u} = \mathbf{u_{h/2}}$ and is approximately

$$Q^2\left(\mathbf{u_{h/2}}\right) \approx 1/8, \tag{4.53}$$

by taking the cube of Eq. (4.39).

The specific form of $S\left(\mathbf{N}, \mathbf{u}\right)$ in Eq. (4.43) assumes that the crystals are all parallelepipeds. In reality the crystals will adopt a variety of different sizes and shapes. This is of no significance in practice however, since $Q^2(\mathbf{u})$ in Eq. (4.50) is the average over the appropriate set of shape transforms $S\left(\mathbf{N}, \mathbf{u}\right)$, and $Q^2(\mathbf{u})$ is estimated from the diffraction data and not calculated using Eq. (4.50), as to be described in the next chapter. Other effects such as crystal disorder are also automatically incorporated into $Q^2(\mathbf{u})$ that is estimated from the diffraction data [Dilanian et al. 2013].

4.5 CONCLUDING REMARKS

It is assumed here that the nanocrystal lattice is well-ordered. It is possible that disorder may be significant as a result of the large surface-to-volume ratio of nanocrystals. However, the fact that experimental diffraction patterns from nanocrystals as small as nine unit cells across show strong interference fringes [Chapman et al. 2011] indicates that the effects of disorder can be quite small, perhaps because the crystals are smaller than their potential mosaic domains.

Disorder that is dominated by uncorrelated distortions of the crystal lattice will decrease the amplitudes of the Bragg reflections with increasing resolution but will not change their profile shapes, and will add a slowly varying diffuse background [Millane and Stroud 1997, Welberry 2004] neither of which will have serious effects on the analyses considered. If correlated distortions of sufficient magnitudes are present, then the Bragg profiles will broaden with increasing resolution, which would require further considerations.

In this chapter we have described the diffraction from a collection of nanocrystals of different sizes and shown that the correlation between their edge lengths and the probability distribution of their sizes has only a weak effect on the resulting averaged

shape transform. The diffraction can be observed between the Bragg peaks which gives rise to the prospect that diffraction data recorded from a stream of nanocrystals in an XFEL can be processed to provide an estimate of the continuous molecular transform, thereby offering the possibility of direct phase retrieval and structure determination in the absence of other experimental data, a prospect which is explored in the next chapter.

Chapter 5

PHASE RETRIEVAL OF NANOCRYSTAL DIFFRACTION WITH NOISE

5.1 INTRODUCTION

A method of retrieving diffraction phases using the averaged diffracted intensity obtained from non-identical nanocrystals of varying sizes would allow XFELs to address the last of the three most important problems in protein X-ray crystallography - the inability to measure X-ray phase.

Iterative phase retrieval methods can in principle be used to reconstruct the protein electron density if the molecular transform is known. Estimation of the molecular transform from averaged nanocrystal diffractions has been proposed and demonstrated by simulation by Spence et al. [2011]. Their strategy allows the finite-size effect, i.e., the averaged shape transform of the nanocrystals to be estimated without any prior knowledge on the crystal size distribution. The averaged shape transform is then used, together with the diffraction data, to estimate the molecular transform. The inherent difficulty with this method however, is that the diffraction signal between the Bragg reflections is weak and so the derived molecular transform is noise-sensitive at those regions, negatively affecting the success of phase retrieval.

In this chapter, the procedure proposed by Spence et al. [2011] is first described and rigorously justified. The signal-to-noise characteristics of the molecular transform estimate that would be used for phasing are then examined and a strategy where the diffracted intensity data is selected to ameliorate the effects of noise is described. Iterative phase retrieval algorithms are subsequently used to reconstruct electron densities from simulated diffraction data.

5.2 DETERMINING THE CONTINUOUS MOLECULAR TRANSFORM

As briefly alluded to in Chapter 3, once the averaged diffracted intensity from all usable nanocrystal diffraction patterns has been assembled, two different pathways for

determining the structure of the unit cell lead off from this point. The first uses the structure factor amplitudes, $|F_{\mathbf{h}}|$, together with phasing methods such as molecular replacement to retrieve the phases. The second pathway results from the crystals in SFX being small in size, i.e., have small numbers of unit cells, which means diffraction between the Bragg reflections can be observed, leading to phase determination methods that is able to utilise the entire continuous molecular transform amplitude, $|F(\mathbf{u})|$. The term "continuous" is used here in the sense that one has access to the diffraction amplitudes between, as well as at, the reciprocal lattice points. If the continuous molecular transform can be obtained, then the phase problem can be solved without any additional experimental data via phase retrieval techniques for single, non-periodic objects, as described in Sec. 3.3.3.5. The single object in this case is the unit cell. A method by which $|F(\mathbf{u})|$ can be determined from the averaged nanocrystal diffraction intensity is now described.

5.2.1 "Dividing-out" the Shape Transform

The continuous magnitude of the molecular transform is required to reconstruct the electron density map of the unit cell through *ab initio* phase retrieval techniques and is within grasp from Eq. (4.34). It is clear from that equation that the magnitude of the molecular transform can be obtained by dividing the averaged measured intensity by the averaged squared shape transform, i.e.,

$$|F(\mathbf{u})|^2 = \frac{\big\langle I_n(\mathbf{u})\big\rangle_n}{\big\langle S_n^2(\mathbf{u})\big\rangle_n}. \tag{5.1}$$

In general, the nanocrystal size distribution $p(\mathbf{N})$, and thence $\langle S_n^2(\mathbf{u})\rangle_n$, is unknown. However, Spence et al. [2011] describe a method for estimating $\langle S_n^2(\mathbf{u})\rangle_n$ directly from the diffraction data, using the idea that the shape transform is periodic and that it is independent of the molecular transform.

The shape transform, as we have seen in the last chapter, is periodic with a period equal to the Bragg spacing. The shape transform function is therefore identical around each Bragg reflection. The molecular transform is the Fourier transform of the contents of the unit cell and is thus uncorrelated with the shape transform. Averaging together the diffraction intensities within fixed regions around many different reflections should therefore preserve the shape transform but encourage the molecular transform to tend towards a mean function. In fact, as is shown next, this mean function is actually a constant, so the process smoothes out the molecular transform to generate an average proportional to the shape transform for that particular diffraction pattern. Thus, averaging together each individual neighbourhood around each Bragg peak in many patterns from crystals of varying sizes gives an estimate of the averaged shape transform around one Bragg peak. The estimate is then one period of $\langle S_n^2(\mathbf{u})\rangle_n$ which can be

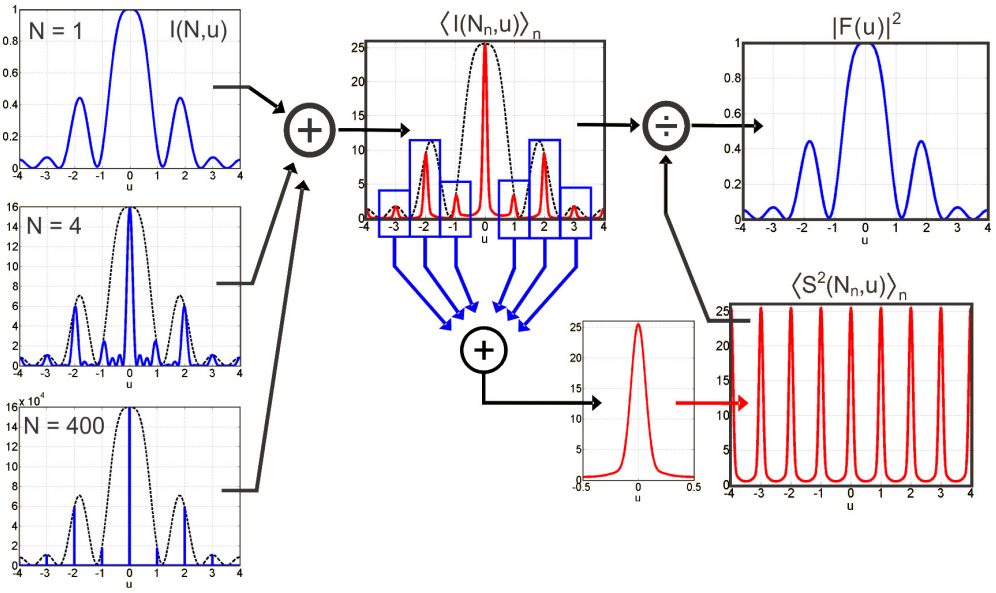

Figure 5.1 The continuous molecular transform recovery process proposed by Spence et al. [2011]. Diffraction by crystals of various sizes are first summed together to give the averaged diffracted intensity. The intensity values within each Bragg volume (blue boxes) of the averaged diffracted intensity are aligned and summed. The output from the summer yields the averaged squared shape transform around one Bragg region which is then replicated to form the averaged shape transform for division to extract the molecular transform.

translated periodically throughout the reciprocal lattice to yield an estimate of the averaged squared shape transform function. The whole process is depicted in Fig. 5.1. Spence et al. [2011] have simulated this process for realistic experimental conditions and shown that it is feasible to recover the molecular transform magnitude through the division of the averaged diffraction intensity by the estimate of the averaged squared shape transform obtained in this way.

5.2.2 Theoretical Justification

The procedure described above can be shown to converge to the averaged squared shape transform as follows. Define a "Bragg volume" associated with each Bragg peak as the region with boundaries equi-distant between that peak and the nearest neighbouring peaks[1]. The diffracted intensity in each Bragg volume is then averaged together with the intensity from all other Bragg volumes. Let $\langle\rangle_{B,\mathbf{h}}$ denote averaging of the Bragg volumes over all reciprocal lattice points, and $\langle\rangle_{B,\mathbf{h},n}$ averaging of the Bragg volumes over all reciprocal lattice points and over all patterns. Averaging together all Bragg volumes from all diffraction patterns then amounts to

$$\left\langle I_n(\mathbf{u}-\mathbf{u_h})\right\rangle_{B,\mathbf{h},n} = \left\langle\, |F(\mathbf{u}-\mathbf{u_h})|^2\, S_n^2(\mathbf{u}-\mathbf{u_h})\right\rangle_{B,\mathbf{h},n}. \tag{5.2}$$

[1]This is known as a *Wigner-Seitz cell* in crystallographic jargon and more generally, a *Voronoi cell* when these volumes are not restricted to centre on uniformly spaced grid points.

Since the shape transform is invariant under translations $\mathbf{u_h}$, we have $S_n^2(\mathbf{u}-\mathbf{u_h}) = S_n^2(\mathbf{u})$, and since the molecular transform is the same for all patterns, it is independent of n. Furthermore, the molecular transform and the shape transform are uncorrelated, so that Eq. (5.2) reduces to

$$\left\langle I_n(\mathbf{u}-\mathbf{u_h})\right\rangle_{B,\mathbf{h},n} = \left\langle \left|F(\mathbf{u}-\mathbf{u_h})\right|^2 \right\rangle_{B,\mathbf{h}} \left\langle S_n^2(\mathbf{u})\right\rangle_{B,n}. \tag{5.3}$$

Consider now the average over the molecular transform, $\left\langle \left|F(\mathbf{u}-\mathbf{u_h})\right|^2 \right\rangle_{B,\mathbf{h}}$, in Eq. (5.3). Working in 1-D for clarity of notation, denote the sum of a total of T shifted versions of $|F(u)|^2$ by $K_T(u)$, where the magnitude of each shift is equal to the distance between the Bragg reflections, $1/a$, so that

$$K_T(u) = \sum_h^{(T)} \left| F\left(u - \frac{h}{a}\right)\right|^2, \tag{5.4}$$

where $\sum_h^{(T)}$ denotes the sum over some T shifts with each shift indexed by h. A particular Bragg volume of $K_T(u)$ that corresponds to a total of T overlapping shifted versions of $|F(u)|^2$ is therefore equal to $\left\langle \left|F(u-u_h)\right|^2 \right\rangle_{B,h}$ after summing T Bragg volumes. Equation (5.4) can be written as

$$K_T(u) = \left|F\left(u\right)\right|^2 \otimes \sum_h^{(T)} \delta\left(u - \frac{h}{a}\right). \tag{5.5}$$

Letting $T \to \infty$ and taking the inverse Fourier transform, Eq. (5.5) becomes

$$k_\infty(x) = aA(x) \sum_{h=-\infty}^{\infty} \delta\left(x - ha\right), \tag{5.6}$$

where $A(x) = f(x) \otimes f(-x)$ is the autocorrelation of the unit cell electron density $f(x)$ and $k_\infty(x)$ is the inverse Fourier transform of $K_\infty(u)$. Since $f(x)$ is zero outside $(-a/2, a/2)$, $A(x)$ is zero outside $(-a, a)$, so that Eq. (5.6) collapses to

$$k_\infty(x) = aA(0)\delta\left(x\right) = a\int_{-\infty}^{\infty} f^2\left(x\right) \mathrm{d}x\, \delta\left(x\right). \tag{5.7}$$

Taking the Fourier transform of Eq. (5.7) yields

$$K_\infty(u) = a\int_{-\infty}^{\infty} f^2\left(x\right) \mathrm{d}x = K, \tag{5.8}$$

i.e. as $T \to \infty$, $K_T(u)$ tends towards the sum of the squared values of the molecular density, a constant which we denote by K. This implies that the average, $\left\langle \left|F(u-u_h)\right|^2 \right\rangle_{B,h}$, tends to the constant K/T for a large number of reflections. Since T is large but finite in practice, this shows that the average over the Bragg volumes of the diffracted

intensities converges to a function approximately proportional to one Bragg volume of the averaged shape transform. Upon replicating and shifting the average to all reciprocal lattice points, a process denoted here for convenience by the operation $\mathcal{R}\{F(\mathbf{u})\} = \sum_{\mathbf{h}} F(\mathbf{u} - \mathbf{u_h})$, we have

$$Q^2(u) \propto \mathcal{R}\big\{\big\langle I_n(u - u_h)\big\rangle_{B,h,n}\big\}, \tag{5.9}$$

as required. The above analysis extends straightforwardly to 3-D. Thus, the average over all Bragg reflections from all patterns generates an estimate of the averaged squared shape transform which can be inserted into Eq. (5.1) to estimate the molecular transform from the measured intensity. An estimate of the molecular transform can thus be obtained from the averaged intensities alone, to within a constant factor, i.e.,

$$|F(\mathbf{u})|^2 \propto \frac{\big\langle I_n(\mathbf{u})\big\rangle_n}{\mathcal{R}\big\{\big\langle I_n(\mathbf{u} - \mathbf{u_h})\big\rangle_{B,\mathbf{h},n}\big\}}. \tag{5.10}$$

5.2.3 Comparison of Two Merging Methods

As mentioned in Sec. 3.5.2, merging the diffracted intensities can also be done with the aim of generating the structure factor amplitudes. The process was discussed in Sec. 3.5.2 and involves averaging together the diffracted intensities from a small volume around each reciprocal lattice point. The structure factor amplitudes $|F_{\mathbf{h}}|$ can be obtained from the averaged diffracted intensity using Eq. (3.27).

Equations (3.27) and (5.10) express $|F_{\mathbf{h}}|$ and $|F(\mathbf{u})|$ respectively in terms of measurable quantities obtainable from SFX experiments. The two equations are similar in that they both use the Monte Carlo integration idea to merge the diffraction data, but Eq. (3.27) yields the discrete structure factors whereas Eq. (5.10) generates the continuous molecular transform. The difference between the two methods lies in the nature of the integration domains around the reciprocal lattice points. The first chooses the domain to be sufficiently small and depends on the shape transform and the molecular transform to not vary significantly within it, the second chooses the domain to cover the entire Bragg volume exactly and depends on the molecular transform to converge to a constant value given a sufficient number of summations. Essentially then, the difference boils down to

$$\langle |S_n(\mathbf{u})|^2 \rangle_{\delta,n} = \text{constant}, \tag{5.11}$$

$$\langle |S_n(\mathbf{u})|^2 \rangle_{B,n} = \text{one period of } \langle |S_n(\mathbf{u})|^2 \rangle. \tag{5.12}$$

Note that $|F_{\mathbf{h}}|$ can also be determined by sampling $|F(\mathbf{u})|$ on the reciprocal lattice points, which may allow independent verification of the merged structure factors.

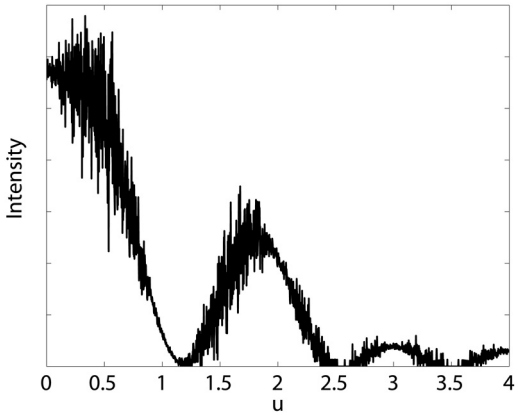

Figure 5.2 Intensity of the molecular transform recovered by dividing a noisy averaged intensity through by the averaged shape transform. The reciprocal space axis is normalised such that the Bragg peaks lie on integer values.

5.3 NOISE AMPLIFICATION

In practice, small values of the averaged shape transform between the Bragg reflections make estimates of the molecular transform in these regions by the division procedure described above noise-sensitive. The sensitivity to noise can be reasoned as follows.

Let the noisy measured and averaged diffraction intensity be additively decomposed into its noiseless component, $\langle I_n(u)\rangle_n$, and the noise contribution, $noise(u)$, then the Fourier magnitude data used for phasing, denoted $|F(u)|_P$, is related to the measured data by

$$\left\langle I_n(u)\right\rangle_n + noise(u) = |F(u)|_P^2\, Q^2(u). \tag{5.13}$$

Rearranging and substituting for $\langle I_n(u)\rangle_n = |F(u)|^2\, Q^2(u)$ from Eq. (4.34) shows that

$$|F(u)|_P^2 = |F(u)|^2 + \frac{noise(u)}{Q^2(u)}, \tag{5.14}$$

so that the measurement noise is amplified in the phasing magnitude by the inverse of the averaged squared shape transform, with the amplification being smallest at the Bragg peaks and largest half-way in between as illustrated in Fig. 5.2. The statistics of the $noise(u)$ function is now quantified.

5.3.1 Noise Statistics for a Large Number of Patterns

The detector pixel positions, for a particular nanocrystal orientation, are mapped into reciprocal space and the diffracted intensities are then re-sampled and binned into those locations for processing. Let the mapped detector location in reciprocal space for the ith detector pixel be denoted by $\mathbf{u}_i$ and consider the statistics of the intensity measurement at that location. Denote the value of the squared amplitude of the molecular transform

at the ith detector pixel by I_i, and the value of I_i modulated by the squared shape transform for the nth nanocrystal at that location by $\mathcal{I}_{in}$, i.e. $\mathcal{I}_{in} = S_{in}^2 I_i$ where $S_{in} = S_n(\mathbf{u}_i)$ is the squared shape transform of the nth nanocrystal at the ith detector pixel. A noisy measurement of $\mathcal{I}_{in}$, denote by $\hat{\mathcal{I}}_{in}$ is subsequently recorded at the detector during the diffraction experiment.

At the data processing stage, the averaged measured intensity at the ith pixel, denoted here by I_{Mi}, is calculated by averaging over the available snapshot diffraction patterns so that

$$I_{Mi} = \frac{1}{m}\sum_{n=1}^{m} \hat{\mathcal{I}}_{in}, \tag{5.15}$$

where m is the number of available diffraction patterns. Note that in general, m may depend on i, but here it is assumed that the number of data samples available for merging are the same for each reciprocal space position after data processing. We denote by I_{Pi} the estimate of I_i that is calculated from the data I_{Mi} via the procedure described in the previous sections and is the quantity used for phasing. In other words, $I_{Pi} = I_{Mi}/\hat{Q}_i^2$, where $\hat{Q}_i^2$ is the estimate of $Q_i^2 = Q^2(\mathbf{u}_i)$, the averaged squared shape transform at pixel i, obtained as described in Sec. 5.2. The goal is to find the statistical distribution of I_{Pi} induced by random fluctuations in detector measurements.

Since the recorded diffraction patterns are very weak, the measurements are assumed to be dominated by photon noise. For such a photon-limited system, the noisy output from the detector is governed by a Poisson distribution with mean and variance equal to the measured quantity, in this case, $S_{in}^2 I_i$, giving

$$\hat{\mathcal{I}}_{in} \sim \mathrm{Po}\left(S_{in}^2 I_i\right). \tag{5.16}$$

Averaging $\hat{\mathcal{I}}_{in}$ over m patterns, the resulting I_{Mi} becomes approximately normally distributed for m large by the central limit theorem under the conditions of Lyapunov as described in Sec. 2.2.4.2. The averaging also means that S_{in}^2 becomes Q_i^2. The $\hat{Q}_i^2$ are estimated by averaging the intensities over the Bragg reflections as well as over the diffraction patterns so that the errors in $\hat{Q}_i^2$ are expected to be smaller than those in I_{Mi}. Thus, we replace $\hat{Q}_i^2$ by Q_i^2, i.e. take $\hat{Q}_i^2$ to be exact. This assumption is justified in the next section. With the above assumptions, it is readily shown using Eq. (5.15) that for large m, the averaged measured intensity at the ith pixel is normally distributed with mean $Q_i^2 I_i$ and variance $Q_i^2 I_i/m$, or

$$I_{Mi} \sim \mathcal{N}\left(Q_i^2 I_i, \frac{1}{m} Q_i^2 I_i\right), \tag{5.17}$$

where $\mathcal{N}(\mu, \sigma^2)$ denotes the normal distribution with mean μ and variance σ^2. In an ideal noiseless system, I_{Mi} is of course equal to $Q_i^2 I_i$, which means that the noise on the measured averaged intensity is normally distributed with a mean of zero and a

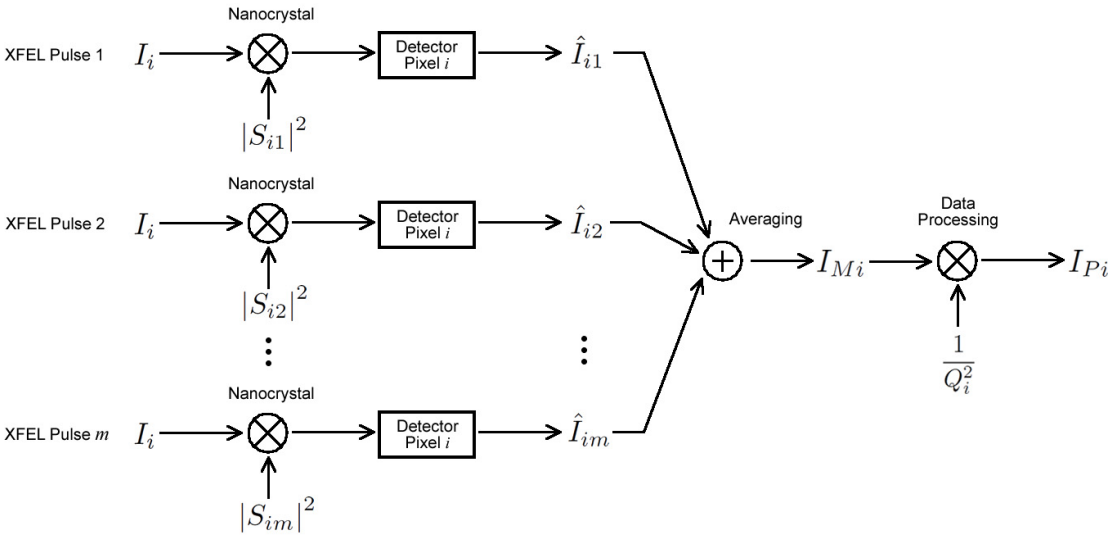

Figure 5.3 System diagram for the process of obtaining the molecular transform from the diffracted intensity for one voxel.

variance of $Q_i^2 I_i/m$ by Eq. (5.17). Since I_{Pi} is obtained by dividing I_{Mi} by Q_i^2, it is also approximately normally distributed and

$$I_{Pi} \sim \mathcal{N}\left(I_i, \frac{I_i}{mQ_i^2}\right), \tag{5.18}$$

so that the noise on the phasing intensity is governed by a zero mean normal distribution with variance $I_i/(mQ_i^2)$. A system diagram illustrating the process described above is shown in Fig. 5.3.

5.3.2 Quantifying the Noise Amplification

Having obtained distributions for I_{Mi} and I_{Pi} along with the distributions for their respective noise terms, the signal-to-noise ratio can be computed. Recall that for a collection of strictly positive measurements, the SNR can be defined as the mean of the signal divided by the standard deviation of the noise (or, equivalently, of the signal). Let SNR_{Mi} and SNR_{Pi} denote the SNR for the measured intensity I_{Mi} and the phasing intensity I_{Pi}, respectively. Using the expressions for the mean and standard deviation of the noise signal derived above, it can be shown that

$$\mathrm{SNR}_{Mi} = \mathrm{SNR}_{Pi} = \sqrt{mQ_i^2 I_i}, \tag{5.19}$$

i.e., the SNRs for individual pixels after the averaging and division steps are the same. The overall SNR of the whole data set (all p voxels) is of interest and is calculated as the mean of the signal over all detector pixels, divided by the square root of the mean of the variances of the noise over all detector pixels, and is denoted SNR_M and SNR_P,

for the measured and phasing intensities, respectively. Using Eqs. (5.17) and (5.18) gives

$$\mathrm{SNR}_M = \sqrt{\sum_{i=1}^{p} Q_i^2 I_i}\sqrt{\frac{m}{p}} \tag{5.20}$$

and

$$\mathrm{SNR}_P = \frac{\sum_{i=1}^{p} I_i}{\sqrt{\sum_{i=1}^{p} I_i/Q_i^2}}\sqrt{\frac{m}{p}}. \tag{5.21}$$

To simplify these expressions, let there be L pixels in each Bragg volume and B Bragg volumes altogether, then $p = LB$. Replacing the indexing variable i with b and l, we can write

$$\sum_{i=1}^{p} Q_i^2 I_i = \sum_{b=1}^{B}\sum_{l=1}^{L} Q_{bl}^2 I_{bl}. \tag{5.22}$$

Concentrating on simplifying the sum in Eq. (5.20) for the moment, the fact that all Bragg regions have identical averaged shape transforms means that

$$\sum_{i=1}^{p} Q_i^2 = B\sum_{l=1}^{L} Q_l^2, \tag{5.23}$$

where the sum over l is over the pixels in one Bragg volume. Similarly, summing the molecular transform intensity over many Bragg reflections converges to a constant, as shown in Sec. 5.2.2, we have

$$\sum_{i=1}^{p} I_i = L\sum_{b=1}^{B} I_b, \tag{5.24}$$

where the sum over b is over the equivalent pixel positions in all Bragg volume. The above suggests that Q_{bl}^2 is independent of b and I_{bl} do not vary very much over l after the sum, allowing us to write

$$\sum_{i=1}^{p} Q_i^2 I_i = \sum_{b=1}^{B}\sum_{l=1}^{L} Q_l^2 I_b.$$

Factoring the double summation and multiplying by LB in the numerator and denominator, we have

$$\sum_{i=1}^{p} Q_i^2 I_i = \frac{1}{LB}\cdot B\sum_{l=1}^{L} Q_l^2 \cdot L\sum_{b=1}^{B} I_b.$$

By Eqs. (5.23) and (5.24) and recalling that $p = LB$, this simplifies to

$$\sum_{i=1}^{p} Q_i^2 I_i = \frac{1}{p}\left(\sum_{i=1}^{p} Q_i^2\right)\left(\sum_{i=1}^{p} I_i\right). \tag{5.25}$$

Substituting this result into Eq. (5.20) for SNR_M and using a similar argument to factor the sum in the denominator of Eq. (5.21) involving the inverse averaged shape transform for SNR_P gives

$$\mathrm{SNR}_M = \frac{1}{p}\sqrt{m\left(\sum_{i=1}^{p} Q_i^2\right)\left(\sum_{i=1}^{p} I_i\right)} \tag{5.26}$$

and

$$\mathrm{SNR}_P = \sqrt{m\left(\frac{1}{\sum_{i=1}^{p} Q_i^{-2}}\right)\left(\sum_{i=1}^{p} I_i\right)}. \tag{5.27}$$

The overall measured signal-to-noise ratio, SNR_M, thus increases as the square root of the total number of patterns m and decreases in proportion to the inverse of the number of detector pixels. This makes sense as the total number of patterns determines the number of different measurements at the same location while the number of detector pixels determines the number of different measurements at different locations.

The two SNRs were calculated for a diffraction simulation in 1-D as a function of mean crystallite size. A Gaussian crystal size distribution with standard deviation equal to one third of the mean as in Sec. 4.3 was used. The sum $\sum_{i=1}^{p} I_i$ in Eqs. (5.26) and (5.27) was set to unity, as was m, as these do not affect the dependence on the mean crystal size. A total of 11 Bragg peaks were simulated and the number of detector pixels $p = 1101$. The results are shown versus mean crystallite size in Fig. 5.4(a).

The behaviours of SNR_M and SNR_P in Fig. 5.4(a) can be explained as follows. SNR_M and SNR_P are proportional to the square root of $\sum_i Q_i^2$ and $1/\sum_i Q_i^{-2}$, respectively. As shown in Sec. 4.3.1, the maximum value of Q_i^2 is proportional to the second moment of the crystal size distribution, which increases without bound as the mean crystallite size increases. This explains the monotonic increase of SNR_M with mean crystallite size. For SNR_P, the increase in SNR for larger crystals, due to stronger scattering, is balanced by the division by a narrower averaged shape transform for the larger crystals, and SNR_P approaches a constant as the mean crystallite size increases. An approximation to SNR_P can be obtained as follows. Using Eq. (4.41), the sum $\sum_{i=1}^{p} Q_i^{-2}$ over a large number of detector pixels is given approximately by $p\langle 2\sin^2(\pi a u)\rangle_u$ which is equal to p. Substituting into Eq. (5.27) gives

$$\mathrm{SNR}_P \approx \sqrt{\frac{m}{p}\sum_{i=1}^{p} I_i}. \tag{5.28}$$

A key factor in obtaining the continuous molecular transform by dividing-out the averaged squared shape transform is the effect on the SNR that results from the division, and how this is influenced by the mean crystallite size. This effect is quantified by calculating the ratio of the SNR at the phasing stage and comparing it to the SNR of

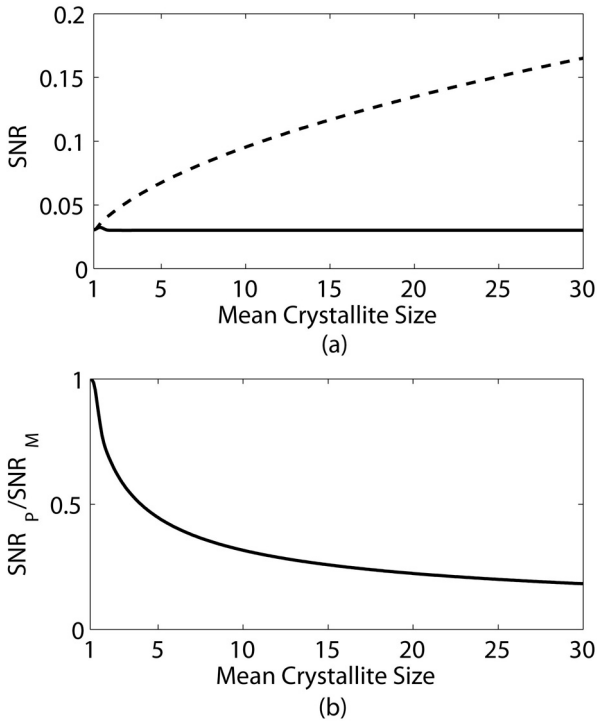

Figure 5.4 (a) Overall SNR of the diffraction pattern measured at the detector, SNR_M (dashed line) and the phasing SNR, SNR_P (solid line), and (b) the ratio $\mathrm{SNR}_P/\mathrm{SNR}_M$ as a function of mean crystal size.

the data measured at the detector, i.e. $\mathrm{SNR}_P/\mathrm{SNR}_M$, which, using Eqs. (5.26) and (5.27), is given by

$$\frac{\mathrm{SNR}_P}{\mathrm{SNR}_M} = p\left(\sum_{i=1}^{p} Q_i^2 \sum_{i=1}^{p} \frac{1}{Q_i^2}\right)^{-\frac{1}{2}}. \tag{5.29}$$

Equation (5.29) can be written as

$$\begin{aligned}\frac{\mathrm{SNR}_P}{\mathrm{SNR}_M} &= \frac{\left(\frac{1}{p}\sum_{i=1}^{p} Q_i^{-2}\right)^{-\frac{1}{2}}}{\left(\frac{1}{p}\sum_{i=1}^{p} Q_i^{2}\right)^{\frac{1}{2}}} \\ &= \frac{M_{-2}(Q_i)}{M_2(Q_i)},\end{aligned} \tag{5.30}$$

where

$$M_\alpha(x_i) = \left(\frac{1}{N}\sum_{i=1}^{N} x_i^\alpha\right)^{\frac{1}{\alpha}} \tag{5.31}$$

is the generalised mean or power mean of exponent α of the collection of positive real numbers $\{x_1, \cdots, x_N\}$. The inequality between two power means states that if two real numbers α and β satisfy $\alpha < \beta$ then $M_\alpha(x_i) \leq M_\beta(x_i)$ with equality if and only if

$x_1 = \cdots = x_N$ [Bullen 2003], leading to the bound

$$\mathrm{SNR}_P \leq \mathrm{SNR}_M. \tag{5.32}$$

In words, Eq. (5.32) says that the SNR after the division process always decreases or at most stays the same compared to the SNR of the measurements before the division.

The SNR ratio of Eq. (5.29) plotted as a function of the mean crystallite size for the same 1-D simulation as above is shown in Fig. 5.4(b). It can be seen that the ratio is unity for crystals with a single unit cell (single molecule), corresponding to the upper bound given in Eq. (5.32) and decreases as the mean crystallite size increases. Note that SNR_P increases with increasing incident X-ray flux and with an increasing number of patterns m. The results presented in this section extend straightforwardly to the three-dimensional case and the signal-to-noise ratios show the same kind of general behaviour as a function of mean crystallite size, although they will be different in detail, depending on the kinds of crystal shapes present.

5.3.3 Errors in the Estimate of $Q^2(\mathbf{u})$

Consider now the statistics of the estimates for the averaged squared shape transform $\hat{Q}_i^2$ at the ith pixel. From Eq. (5.3),

$$\big\langle I_n(\mathbf{u}-\mathbf{u_h})\big\rangle_{B,\mathbf{h},n} = \big\langle\, |F(\mathbf{u}-\mathbf{u_h})|^2 \,\big\rangle_{B,\mathbf{h}} Q^2(\mathbf{u}), \tag{5.33}$$

and the results in Sec. 5.2.2 show that

$$\big\langle\, |F(\mathbf{u}-\mathbf{u_h})|^2 \,\big\rangle_{B,\mathbf{h}} = \frac{K}{T}, \tag{5.34}$$

where K is a constant proportional to the sum of the squared electron density inside the unit cell and T is the total number of Bragg volumes summed. Therefore,

$$\begin{aligned}
\hat{Q}_i^2 &= \frac{T}{K}\big\langle I_n(\mathbf{u}_i-\mathbf{u_h})\big\rangle_{B,n} \\
&= \frac{T}{K}\big\langle I_M(\mathbf{u}_i-\mathbf{u_h})\big\rangle_{B} \\
&= \frac{1}{K}\sum_{\mathbf{h}}^{(T)} I_M(\mathbf{u}_i-\mathbf{u_h}).
\end{aligned} \tag{5.35}$$

Furthermore,

$$K = \sum_{\mathbf{h}}^{(T)} I(\mathbf{u}-\mathbf{u_h}), \tag{5.36}$$

and the mean of I_{Mi} is $Q_i^2 I_i$. The expectation of $\hat{Q}_i^2$ is then

$$\begin{aligned}
E\Big[\hat{Q}_i^2\Big] &= E\Big[\frac{1}{K}\sum_{\mathbf{h}}^{(T)} I_M(\mathbf{u}_i - \mathbf{u}_\mathbf{h})\Big] \\
&= \frac{1}{K}\sum_{\mathbf{h}}^{(T)} E\Big[I_M(\mathbf{u}_i - \mathbf{u}_\mathbf{h})\Big] \\
&= \frac{1}{K}\sum_{\mathbf{h}}^{(T)} Q_{i-\mathbf{h}}^2 I_{i-\mathbf{h}} \\
&= \frac{1}{K}\sum_{\mathbf{h}}^{(T)} Q_i^2 I_{i-\mathbf{h}} \qquad \because\ Q_i^2 \text{ is the same for all Bragg regions} \\
&= \frac{1}{K} Q_i^2 \sum_{\mathbf{h}}^{(T)} I_{i-\mathbf{h}} \\
&= \frac{1}{\sum_{\mathbf{h}}^{(T)} I(\mathbf{u} - \mathbf{u}_\mathbf{h})}\, Q_i^2 \sum_{\mathbf{h}}^{(T)} I(\mathbf{u}_i - \mathbf{u}_\mathbf{h}) \\
&= Q_i^2. \qquad (5.37)
\end{aligned}$$

The last equality comes about because K is independent of $\mathbf{u}$ and so $K(\mathbf{u}) = K(\mathbf{u}_i) = K$ for all i.

The variance of I_{Mi} is $Q_i^2 I_i/m$ as derived from Sec. 5.3.1. The variance of $\hat{Q}_i^2$ is then

$$\begin{aligned}
Var\Big[\hat{Q}_i^2\Big] &= Var\Big[\frac{1}{K}\sum_{\mathbf{h}}^{(T)} I_M(\mathbf{u}_i - \mathbf{u}_\mathbf{h})\Big] \\
&= \frac{1}{K^2}\sum_{\mathbf{h}}^{(T)} Var\Big[I_M(\mathbf{u}_i - \mathbf{u}_\mathbf{h})\Big] \\
&= \frac{1}{K^2}\sum_{\mathbf{h}}^{(T)} \frac{1}{m} Q_{i-\mathbf{h}}^2 I_{i-\mathbf{h}} \\
&= \frac{1}{K^2}\frac{1}{m} Q_i^2 \sum_{\mathbf{h}}^{(T)} I_{i-\mathbf{h}} \\
&= \frac{1}{\Big(\sum_{\mathbf{h}}^{(T)} I(\mathbf{u} - \mathbf{u}_\mathbf{h})\Big)^2}\, \frac{1}{m} Q_i^2 \sum_{\mathbf{h}}^{(T)} I(\mathbf{u}_i - \mathbf{u}_\mathbf{h}) \\
&= \frac{1}{\sum_{\mathbf{h}}^{(T)} I(\mathbf{u} - \mathbf{u}_\mathbf{h})}\, \frac{1}{m} Q_i^2 \\
&= \frac{1}{mK} Q_i^2. \qquad (5.38)
\end{aligned}$$

Therefore $\hat{Q}_i^2$ is distributed as

$$\hat{Q}_i^2 \sim \mathcal{N}\left(Q_i^2, \frac{1}{mK}Q_i^2\right). \tag{5.39}$$

Referring to Eqs. (5.17) and (5.39), the errors on the estimates $\hat{Q}_i^2$ are thus smaller than the errors in I_{Mi} by a factor of KI_i, or in symbols,

$$\sigma^2\big(\hat{Q}_i^2\big) = \frac{1}{KI_i}\sigma^2\big(I_{Mi}\big). \tag{5.40}$$

Thus, treating $\hat{Q}_i^2$ as being exact is not too farfetched, provided KI_i is large, which may suggest that the errors in the estimate of the averaged squared shape transform cannot be ignored near the edges of the detector as the diffracted intensity of the molecule, I_i, falls off in reciprocal space.

Note that KI_i has units of intensity unit squared. $\sigma^2\big(I_{Mi}\big)$ also has units of intensity unit squared. Thus $\sigma^2\big(\hat{Q}_i^2\big)$ is unitless as required.

5.4 PHASING IN THE PRESENCE OF NOISE

A key observation from the noise analysis given above is that the SNR for the phasing intensity at a sample position $\mathbf{u}_i$ in reciprocal space is proportional to the value of the averaged shape transform at that position, i.e.

$$\mathrm{SNR}_{Pi} \propto Q(\mathbf{u}_i), \tag{5.41}$$

where SNR_{Pi} is the SNR for the phasing intensity at position $\mathbf{u}_i$, calculated as the mean of the diffracted signal divided by the standard deviation of the noise. Since $Q(\mathbf{u}_i)$ has a wide dynamic range, so does SNR_{Pi}, which needs to be considered in the phase retrieval process. Furthermore, the averaged shape transform $Q^2(\mathbf{u})$, which is estimated from the data, gives an estimate of SNR_{Pi} for each datum. This information can be used to ameliorate the deleterious effects of the variable SNR on the reconstructed electron density.

5.4.1 Selective Oversampling

Since the SNR is spatially variable in reciprocal space, a sensible strategy is to sample the data in such a way as to maximise the SNR of the data needed for phase retrieval. It is necessary to consider both the number of samples used (to ensure that the problem remains well-determined) and their positions in reciprocal space (to maximise the phasing SNR). In data obtained from SFX experiments, the diffraction amplitudes can be sampled onto a grid in reciprocal space that is finer than the reciprocal lattice. If

this grid oversamples reciprocal space by a factor s in each direction relative to the reciprocal lattice, then the oversampling factor O of the 3-D data set is defined as

$$O = s^3. \tag{5.42}$$

Since the continuous molecular transform is used to estimate the electron density, the situation effectively becomes a phase retrieval problem for a single particle (the contents of one unit cell). Under these circumstances, and under the assumption that the molecular support region is approximately convex and centrosymmetric, the minimum number of amplitude data required to uniquely define the electron density is twice the number of Bragg samples at the particular resolution of the data, as described in Sec. 3.3.3.5. In terms of the factor of fine sampling defined in Eq. (5.42), uniqueness requires $O > 2$. The selective sampling approach is then to retain a subset of the oversampled amplitude data with the largest SNR and use these for phase retrieval. The amplitudes at the remaining data points are allowed to float and left to be determined by the constraints in real space. If, as a result of removing data with a low SNR, a proportion $0 \leq P \leq 1$ of the data are retained, then the oversampling factor of the data used for phasing, denoted O_P, is

$$O_P = PO = Ps^3. \tag{5.43}$$

Uniqueness of the solution then requires that $O_P > 2$, and a margin on this inequality is desirable in practice.

Increasing the directional oversampling factor s increases the size of the computational grid by a factor s^3 so it is desirable to keep s as small as possible. Since $Q^2(\mathbf{u})$ provides an estimate of the SNR for each datum, our objective is to retain data points where $Q^2(\mathbf{u})$ is large. For $s = 2$, $O = 8$, assuming a regular sampling grid. But the additional data (i.e. those in addition to the Bragg samples) are midway between the Bragg reflections where $Q^2(\mathbf{u})$ is smallest. This choice of s is therefore unsuitable. For $s = 3$, $O = 27$, the points where $Q^2(\mathbf{u})$ is smallest are avoided, and O is large enough that a significant proportion of the low SNR data can be removed while still satisfying $O_P > 2$. For $s = 4$, $O = 64$, and some of the new samples fall where $Q^2(\mathbf{u})$ is smallest. Given the additional computational cost for $s = 4$, this value does not appear to offer any advantages over $s = 3$. Note that although increasing s increases the number of data, the data points become closer together and therefore the data become more correlated and less information is added. In fact, for the noiseless case, no information is added beyond $s = 2$ as that already satisfies the Shanon sampling criterion.

Since the SNR is proportional to $Q^2(\mathbf{u})$, a sensible approach is to set a threshold, denoted here by ξ, on $Q^2(\mathbf{u})$ and to use the data for which $Q^2(\mathbf{u})$ is greater than ξ. The threshold determines the proportion P of the data that are retained and must be chosen such that O_P is greater than 2. The effect of the threshold is illustrated in 1-D

in Fig. 5.5. The samples that are not used as data (grey regions in Fig. 5.5) are treated as missing data and their values are made to float during the phase retrieval process. The threshold is normalised such that it is unity at the Bragg peaks.

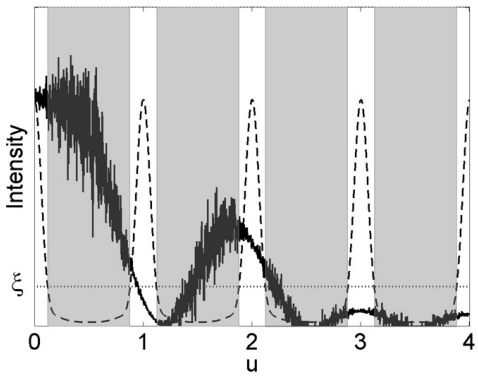

Figure 5.5 A threshold (dotted line) on the averaged shape transform (dashed line) determines the regions (grey) where the intensity data are not used.

For $s = 3$ and the case of an orthorhombic unit cell with a crystallite size distribution having the same marginal probability density in the three directions, there are four possible sampling schemes corresponding to four values of ξ as a result of the symmetry of $Q^2(\mathbf{u})$ and of the reciprocal lattice. These sampling schemes are labelled A, B, C and D in order of increasing ξ and their oversampling factors are listed in Table 5.1. Scheme A corresponds to maximal oversampling (no data removed) and scheme D corresponds to no oversampling (Bragg samples only). Scheme C includes the Bragg samples and the 6 samples closest to the Bragg samples on lines parallel to the reciprocal space axes. Scheme B excludes the "body diagonal" samples only, from the full set of oversampled data of scheme A. Note that in general $Q^2(\mathbf{u})$ will tend to decrease monotonically with distance from the Bragg samples so that in practice it may likely be sufficient, and more convenient, to select sampling schemes based on distance from the reciprocal lattice points rather than on a threshold on the measured $Q^2(\mathbf{u})$.

Table 5.1 Oversampling factor O_P and the proportion of data retained P, for the four sampling schemes for $s = 3$ ($O = 27$).

Sampling Scheme	O_P	P
A	27	1.00
B	19	0.70
C	7	0.26
D	1	0.04

Figure 5.6 shows schematically this idea. Each grid point represents an intensity sample point, the open circles are the samples at the Bragg reflections, which are always kept, closed circles denote samples kept, and no circle represent samples that are not used. By default, the whole set of intensity samples would be utilised during phase

retrieval as in sampling scheme A, but since the samples midway between the Bragg peaks and their immediate surrounding samples are known to be highly noisy, they may be discarded to give rise to the different sampling ratios documented in Table 5.1. In total, there are two unique sample selection schemes, corresponding to unique distances, in 2-D and four unique schemes in 3-D under the regime of three times oversampling in each direction with an orthogonal lattice.

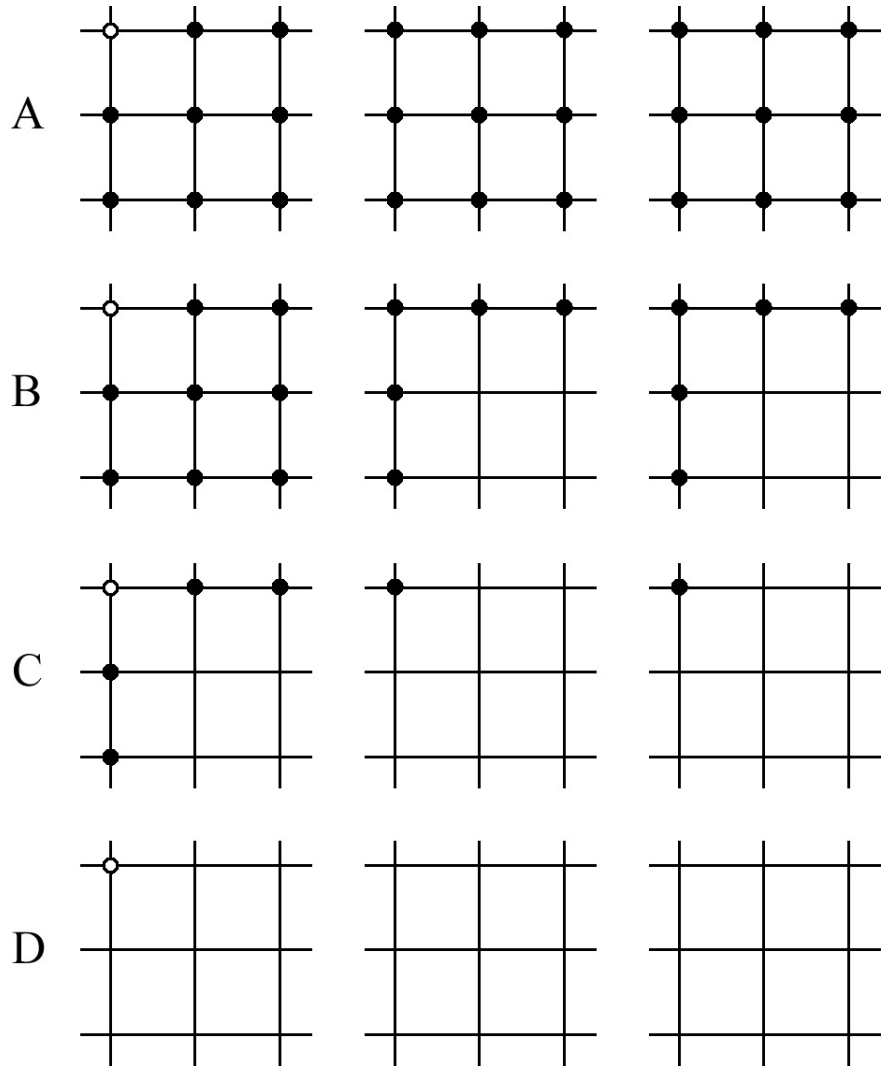

Figure 5.6 Selection schemes for a three times oversampled reciprocal lattice for a 3-D cubic crystal. Grid points are the sampling positions, closed circles denote samples kept and the open circles are the Bragg samples which are always kept. The rows are the different sampling schemes and the columns the reciprocal space direction in the third dimension.

5.5 SIMULATION RESULTS

The selective sampling approach described in the previous section was tested by implementing the different sampling schemes and retrieving the phases of simulated diffraction data using the DM algorithm. The simulations were conducted in 3-D. The molecule used for the simulations was the membrane protein Aquaporin 1 (AQP1) with the electron density data provided by A. K. Mitra from the University of Auckland through the work of Ren et al. [2000]. A $32 \times 32 \times 45$Å section of the electron density map sampled on a 1Å grid was used and is shown in the third column of Fig. 5.9(a) and (b). Reciprocal space was oversampled by a factor of three in each direction. This was done by zero-padding the real space volume out to $96 \times 96 \times 135$ grid points prior to calculating the discrete Fourier transform. The true diffracted intensities were calculated and corrupted with Poisson noise. The noise level on the simulated data was manipulated as follows. Recalling that $\mathrm{Po}(\lambda)$ denotes the Poisson distribution with parameter λ, let I be the original noiseless value of the intensity, k a scaling factor that controls the noise level, and I' drawn from the distribution $\mathrm{Po}(kI)$. The corrupted intensity is then calculated as

$$I_{\text{noisy}} = \frac{1}{k} I'. \tag{5.44}$$

Thus the larger the value of k, the smaller the Poisson noise and the larger the SNR. The SNR of a Poisson random process is the square root of the mean so that in this case the SNR of the sample intensity is $\sqrt{kI}$. The noiseless case can be thought of as having a scaling factor k of infinity.

A Gaussian crystal size distribution with the same mean number of unit cells, denoted μ_N, and a standard deviation $\sigma_N = \mu_N/3$ is imposed on the crystal length in each direction. This ensures that the probability of one side of a crystal being less than one unit cell is negligible and the distribution is truncated at $N = 1$. The number of unit cells in each of the three directions are assumed to be independent. For each value of μ_N considered, the averaged shape transform $Q^2(\mathbf{u})$ was calculated using Eq. (4.50). The noisy amplitudes computed as described above were divided through by this averaged shape transform to calculate the amplitudes to be used for phasing. Intensity samples within a sphere of radius of 5 grid points centred at the origin of reciprocal space were discarded to simulate the effect of missing data due to the gap built into the detector for the undiffracted XFEL beam. The intensity samples outside a sphere of radius 0.5Å^{-1} in reciprocal space were also not used to simulate the resolution sphere, resulting in a resolution of the reconstructed electron density map of 2Å. The resolution limit reduces the number of data points, and hence the oversampling factor, by a factor of $6/\pi \approx 1.9$, i.e., the ratio of the volume of a cube to the volume of a sphere filling that cube.

The DM algorithm was used to reconstruct the electron density map for different average crystal sizes and noise levels. The DM algorithm parameter β was set to 0.7 with $\gamma_A = -1/\beta$ and $\gamma_B = 1/\beta$. Constraints imposed in real space are a tight support and the reality of the electron density. The constraint in reciprocal space is that the Fourier amplitudes are equal to the data values except at the following reciprocal space positions: (1) outside the resolution sphere, (2) around the origin of reciprocal space within the missing zone due to the central beam, and (3) those positions excluded by the particular sampling scheme used. The Fourier amplitudes at all of these missing data regions are allowed to float during the execution of the iterative phase retrieval algorithm.

Progress of the algorithm is monitored by calculating the R-factor, a common measure of error in the Fourier magnitudes of the reconstructed structures in X-ray crystallography (see for example Drenth [2007] for details), in the form of

$$R = \frac{\sum_{\mathbf{u}} \left| |\hat{F}(\mathbf{u})| - |F(\mathbf{u})|_P \right|}{\sum_{\mathbf{u}} |F(\mathbf{u})|_P} \tag{5.45}$$

as a function of iteration, where $|\hat{F}(\mathbf{u})|$ is the Fourier magnitude of the estimated solution $\hat{f}(\mathbf{x})$ at the current iteration obtained using the DM solution projection given by Eq. (2.101), and $|F(\mathbf{u})|_P$ is the Fourier magnitude of the available data. The quality of the reconstruction is measured by calculating the root-mean squared (RMS) error in the reconstructed electron density

$$e = \frac{\|\hat{\mathbf{f}} - \mathbf{f}\|}{\|\mathbf{f}\|} = \sqrt{\frac{\sum_{\mathbf{x}} \left(\hat{f}(\mathbf{x}) - f(\mathbf{x}) \right)^2}{\sum_{\mathbf{x}} f^2(\mathbf{x})}} \tag{5.46}$$

as a function of iteration, where $\|\cdot\|$ is the Euclidean norm, and $f(\mathbf{x})$ and $\mathbf{f}$ denote the true electron density as it appears in real space and collected as a vector, respectively. The same relation applies to $\hat{f}(\mathbf{x})$ and $\hat{\mathbf{f}}$.

The oversampling factor, O_P, and the expected SNR ratio, $\text{SNR}_P/\text{SNR}_M$, for μ between 1 and 25 were calculated using Eq. (5.29) for the four sampling schemes A-D described in the previous section and shown in Fig. 5.7(a). Note that the values of O_P obtained are smaller than those given in Table 5.1 as a result of the removal of the high resolution data outside the resolution sphere as described above. Inspection of Fig. 5.7(a) shows that the best SNR with sufficient oversampling factor ($O > 2$) is obtained with sampling scheme C. The overall SNR ratio for the four sampling schemes is shown versus mean crystal size in Fig. 5.7(b). This shows the deterioration in SNR_P relative to SNR_M with increasing crystal size as noted previously in Sec. 5.3.2 and also the improvement in SNR_P for sampling schemes B and C over using all the samples (scheme A). Note that for a fixed incident X-ray pulse flux, SNR_M will increase with increasing crystallite size.

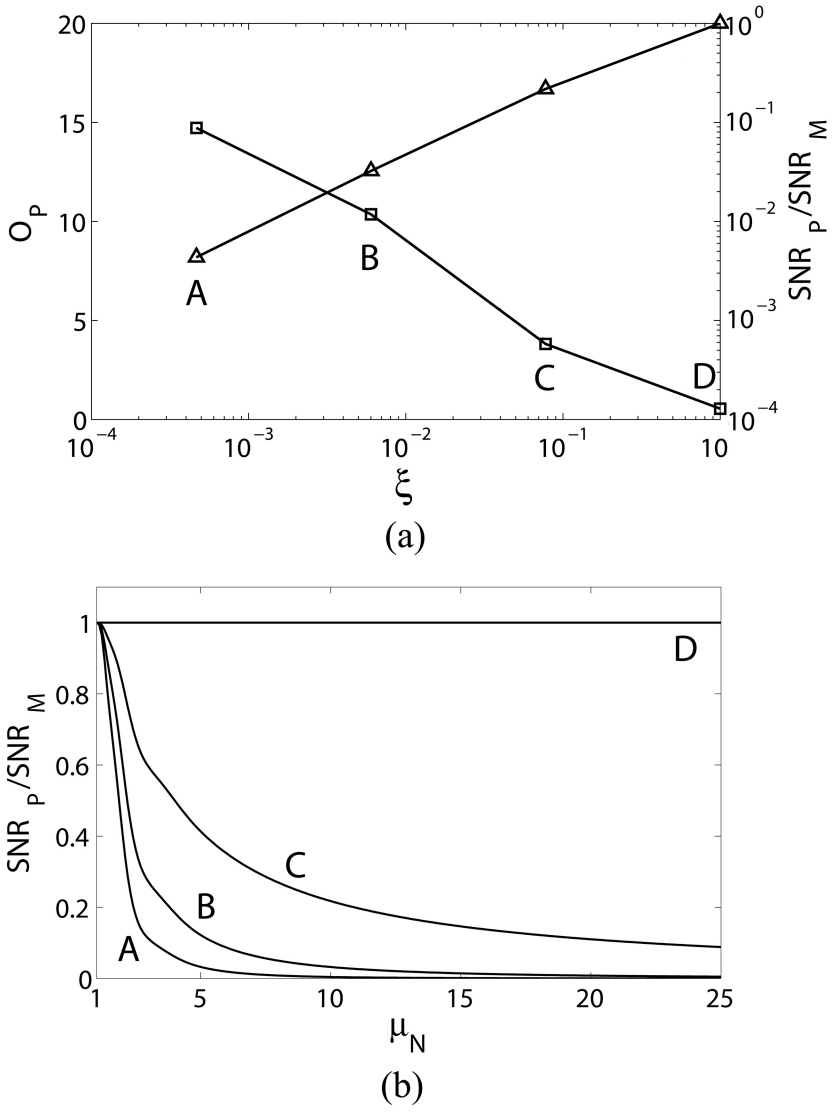

Figure 5.7 (a) The oversampling factor O_P (squares) and the SNR ratio (triangles), for sampling schemes A, B, C and D, with $\mu_N = 10$. (b) The SNR ratio for the four sampling schemes as a function of mean crystallite size.

Simulations were conducted for two SNRs (SNR_M) of the simulated measured intensities, one being $\text{SNR}_M = 100$ and the other $\text{SNR}_M = 20$. A mean crystal size $\mu_N = 10$ was used, and phase retrieval was carried out for the four sampling schemes listed in Table 5.1. The DM algorithm was run for 1000 iterations, starting with a random electron density. The final reconstructed electron density is that with the minimum value of R out of the one thousand iterations.

The R-factor and RMS error versus iteration of the DM algorithm for sampling schemes A and C are shown in Fig. 5.8 for the simulated reconstructions. For both SNRs, the algorithm converges to a good solution for the selective sampling scheme C but not when all samples are used (sampling scheme A). This shows the advantage of selecting the data with the best SNR. The resulting reconstructed electron densities are shown in Fig. 5.9 and it is clear that selective sampling leads to an interpretable density whereas using all the data does not.

The R-factor and electron density RMS error corresponding to the final reconstructions are plotted against the threshold ξ in Fig. 5.10, indicating the effect of the sampling schemes A to D where it can be seen that sampling scheme C gives the best reconstruction. Referring to Figs. 5.7(a) and 5.10, the reconstructions improve as the SNR ratio increases and data are removed until the point at which there is not enough data, at sampling scheme D, where $O < 2$. Note that for sampling scheme D a small R-factor is obtained since the problem is under-constrained and the algorithm can easily find one of a multitude of incorrect solutions that satisfy the data.

For a fixed SNR_M, SNR_P deteriorates as the mean crystal size of the nanocrystals increases. Simulations were conducted for a range of noise levels in the data and a range of crystallite sizes. The resultant RMS errors are displayed as a contour plot in Fig. 5.11 which shows the error in the final reconstructions versus the noise level in the data ($\text{NSR}_M = 1/\text{SNR}_M$) and the mean crystal size, for sampling schemes A and C. Inspection of Fig. 5.11 shows that for a particular crystal size, significantly larger noise levels can be tolerated by using the sample selection method.

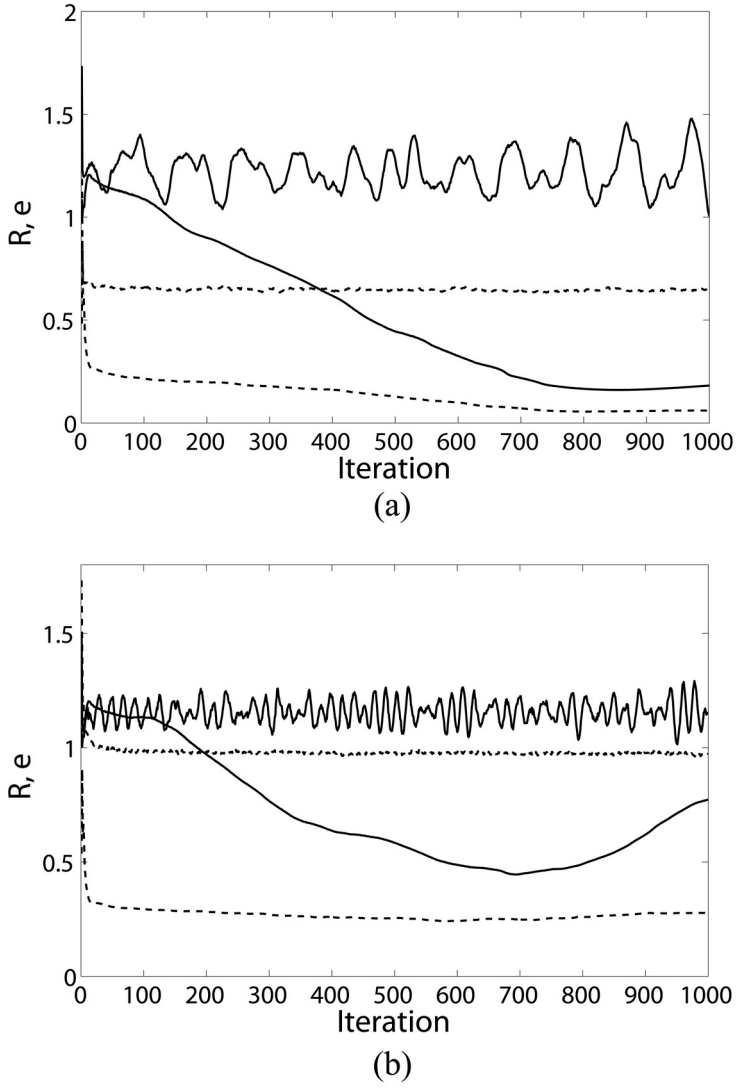

Figure 5.8 R-factor (dotted line) and RMS error (solid line) in the electron density versus iteration using sampling schemes A (upper curves at the 1000th iteration) and C (lower curves at the 1000th iteration), for (a) $\mathrm{SNR}_M = 100$ and (b) $\mathrm{SNR}_M = 20$.

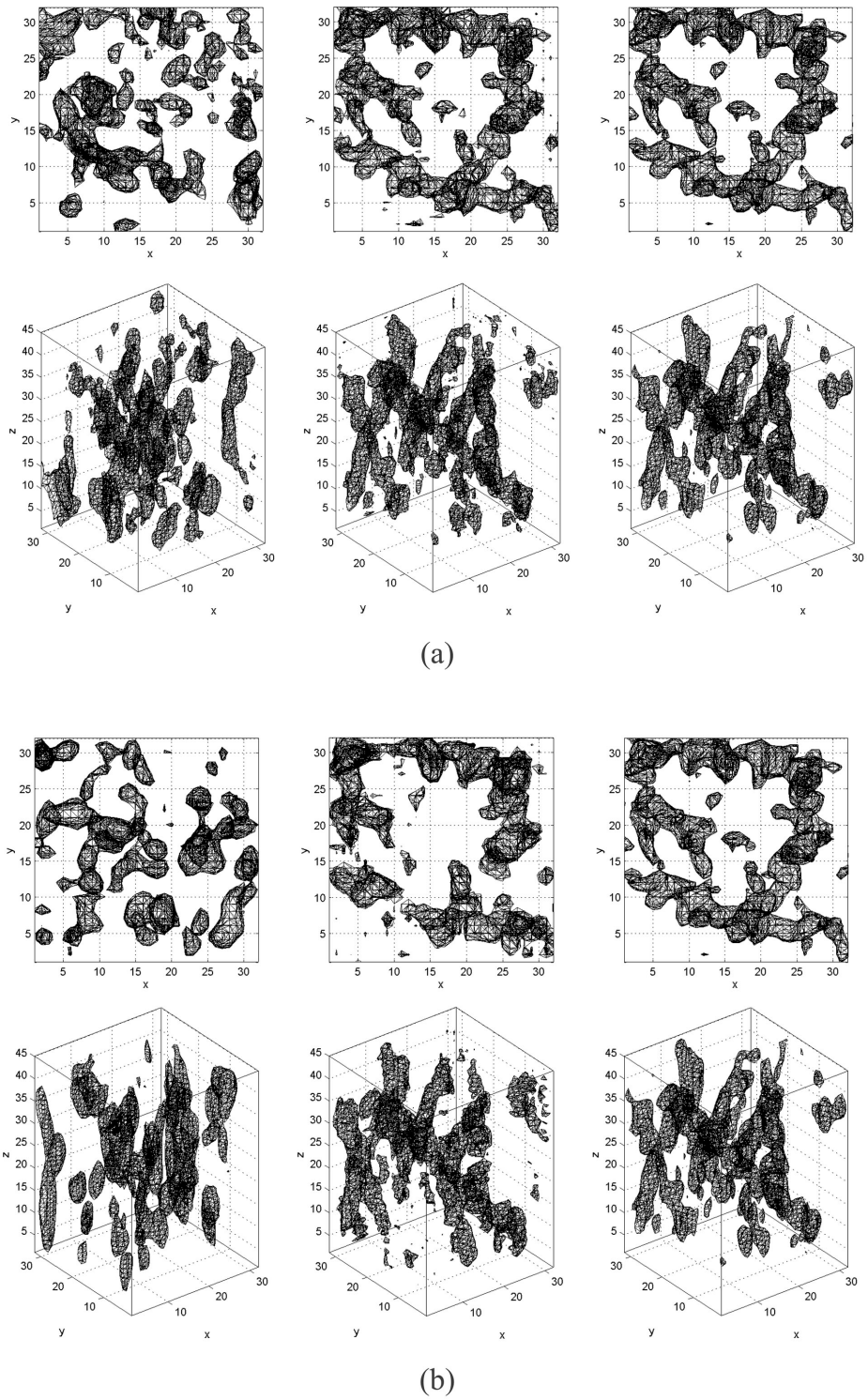

Figure 5.9 Reconstructed electron densities for (a) $\mathrm{SNR}_M = 100$ and (b) $\mathrm{SNR}_M = 20$. The top rows show the projected views along the z-axis and the bottom rows the 3D volume. The left and centre columns are reconstructions using sampling schemes A and C, respectively, and the right column is the true electron density.

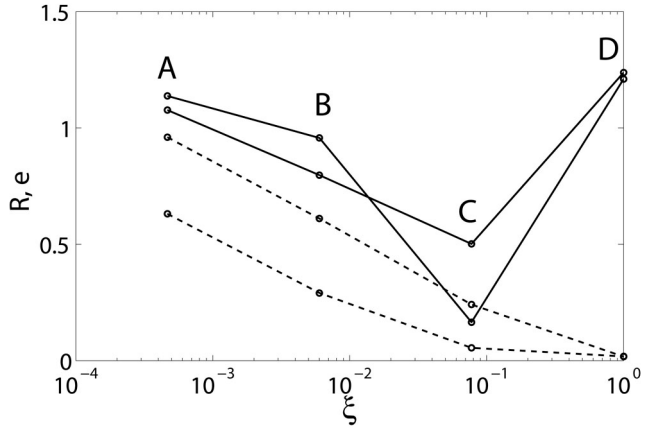

Figure 5.10 R-factor (dotted line) and electron density RMS error (solid line) of the final reconstructions for sampling schemes A-D for $\mathrm{SNR}_M = 100$ (lower curves at C) and $\mathrm{SNR}_M = 20$ (upper curves at C).

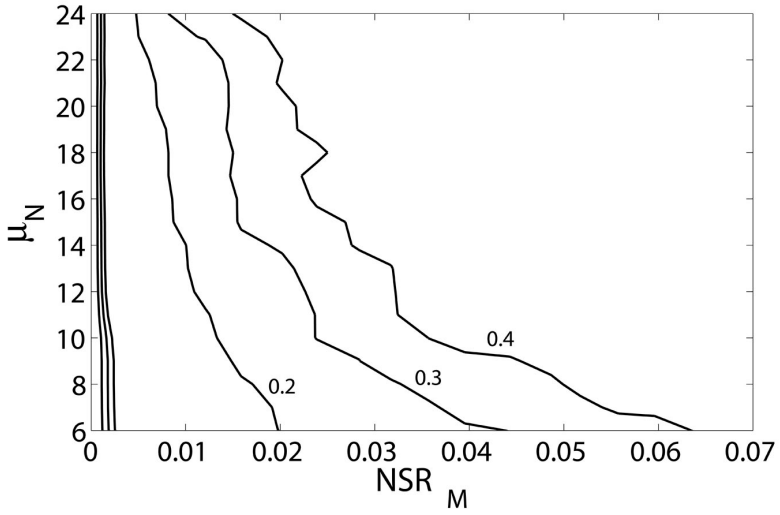

Figure 5.11 Contour plots of constant RMS error in the electron density for sampling scheme A (left three curves) and sampling scheme C (right three curves), versus mean crystal size and noise-to-signal ratio for the measured data NSR_M. The RMS error is contoured at 0.2, 0.3 and 0.4 for both sampling schemes.

5.6 CONCLUDING REMARKS

Nanocrystallography using XFELs offers the possibility of direct phase retrieval courtesy of measurable diffraction data between the Bragg reflections that can be utilised to estimate the molecular transform. Averaging over the diffracted intensity around each Bragg peak for all usable nanocrystal diffraction patterns converges to one period of the averaged squared shape transform. This averaged squared shape transform estimate can then be used to calculate the molecular transform from the diffraction data via a division process first documented by Spence et al. [2011]. The consequence of this approach is that the resulting samples of the molecular transform possess widely varying SNRs at different positions in reciprocal space. The estimated molecular transform has low SNRs at positions midway between the Bragg reflections because the value of the averaged squared shape transform is small at those regions.

To address this problem, a selective sampling strategy that retains only the measured intensity samples that have the highest signal-to-noise ratio is employed. Values of the averaged squared shape transform that are estimated from the diffraction data are used to determine an appropriate sampling scheme, although in practice a scheme based on the distance of samples from the Bragg reflections is likely sufficient. By judiciously selecting samples of the molecular transform to be used in phase retrieval, the effects of noise can be ameliorated and the excluded molecular diffraction amplitudes can be filled-in by the iterative phase retrieval algorithm to allow structure determination at low signal-to-noise ratios.

In terms of simulation, oversampling the reciprocal lattice by a factor of three in each direction allows the potential removal of samples of low SNR while retaining sufficient data for a unique solution and minimising the computational load. Simulations show that selective sampling of the data before application of the phase retrieval algorithm allows reconstruction at lower SNR than if all the data are used. The results show the trade-off between noise levels and crystallite sizes that can be tolerated for direct phasing in nanocrystallography.

The overall SNR of the whole derived data set, SNR_P, is smaller than the overall SNR of the measured data set, SNR_M, as a result of the division by the averaged squared shape transform. This deterioration of the SNR of the phasing amplitude relative to the measured amplitude can be quantified by the ratio given in Eq. (5.29) and worsens (gets smaller) as the mean crystallite size increases. The decrease in $\mathrm{SNR}_P/\mathrm{SNR}_M$ with increasing mean crystallite size gives a quantitative result that may assist in the design of SFX experiments.

Chapter 6

DIFFRACTION CHARACTERISTICS OF NANOCRYSTALS - INCOMPLETE UNIT CELLS

6.1 INTRODUCTION

As described in Chapter 5, with small crystals, there is the potential for using the recorded diffraction at, and between, the Bragg peaks to estimate the diffracted amplitude by a single unit cell. The problem of deducing the electron density of the unit cell from the recorded diffraction then becomes analogous to that of reconstructing a single, non-periodic object (the unit cell) from the amplitude of its Fourier transform. The solution to this phase problem is in priciple unique (up to trivial ambiguities as mentioned in Sec. 3.3.3.5) without any ancillary experimental data, and can be obtained using iterative phase retrieval algorithms. This approach for imaging the unit cell requires a good model of nanocrystals and their continuous diffraction. Up until now, the work presented in the previous chapters had assumed a single kind of unit cell. However, if the unit cell encompasses more than one molecule, which is often the case for protein crystals, then it is possible to define unit cells with different contents that are still able to describe the same crystal.

Redefinition of the unit cell is valid because molecules bind together to form a crystal without regard to the artificial construct of a "unit cell." An aggregate of molecules is not restricted to crystallise into a structure composed entirely of a single kind of full unit cell. The surface of the crystal can have molecules that are situated in any configurations permissible by the symmetry operations between a molecule and its neighbours. Thus, a crystal in the real world will be terminated with molecules that do not necessarily make up a full unit cell. The configurations of molecules that do not form complete unit cells are referred to here as *incomplete* or *partial unit cells*.

An example of a valid redefinition of the unit cell in 3-D can be seen by comparing the orange and purple cells in Fig. 6.1(b). The boundary of the unit cell is modified such that it encompasses different spatial combinations of adjacent molecules in the crystal. This corresponds in the simplest case to shifting the bounding box of the unit cell so that the contents of different unit cells are translationally wrapped versions of

each other. Although these characteristics certainly occur with macroscopic crystals also, as illustrated in Fig. 6.1(a), the effects on the diffraction for large crystals are vanishingly small. Furthermore, the issue of multiple unit cell definitions are non-existent for conventional X-ray crystallography because only the Bragg reflections are of interest. The value of the diffracted intensity at the Bragg peaks does not depend on the unit cell used to describe an infinite crystal, just as redefining the period of a periodic function do not change its Fourier coefficients. It is only when the crystal is small and the diffraction between the Bragg peaks are of interest - both the case for the method of SFX - that the effects of different unit cell definitions need to be considered.

With the possibility of multiple unit cells, the averaged diffraction from an ensemble of crystals will not be simply related to the diffraction from one unit cell [Chen and Millane 2013, Elser 2013, Liu et al. 2014, Kirian et al. 2014]. Modelling the continuous diffraction by nanocrystals in the presence of incomplete unit cells is the topic of this chapter. The terms “molecule” and “asymmetric unit” are used here interchangeably.

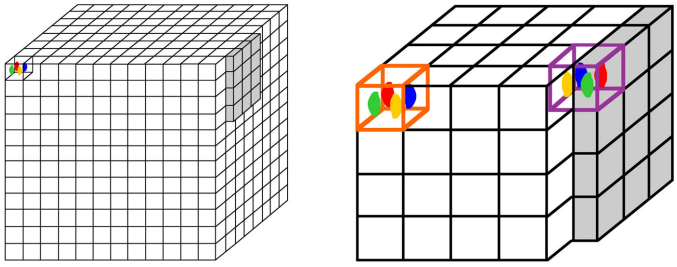

Figure 6.1 Illustration of the edge effect on large and small crystals. (a) A relatively large crystal. (b) A smaller crystal. The grey region indicates partial unit cells, and the coloured blobs denote the four molecules in one unit cell. Different unit cells can be defined in the presence of partial unit cells, as highlighted by the orange and purple bounding boxes in (b).

6.2 NANOCRYSTALS IN ONE-DIMENSION

In 1-D, a crystal can only accrue incomplete unit cells on either of its two ends. Incomplete unit cells are defined in this case as non-overlapping, fractional segments of the full unit cell. The simplest case is when the unit cell is composed of two such segments. Furthermore, there are only two space groups for a 1-D crystal: P1 and P1m. P1 has no symmetry while P1m has a mirror symmetry in the unit cell. A model of 1-D nanocrystals is presented first and the effects of incomplete unit cells on the diffraction and on the estimation of the molecular transform via the process described in Chapter 5 are then explored through simulation.

6.2.1 Unit Cell Definition

Consider a 1-D crystal with two molecules labelled as A and B in the unit cell. Write the electron density of each of these two molecules as $f_A(x)$ and $f_B(x)$. Note that since the only non-trivial space group in 1-D is P1m, $f_A(x) = f_B(-x)$. The two possible definitions of the unit cell in this case are obtained by switching the order of the two molecules as shown in Fig. 6.2. Letting a be the width of the unit cell, the density of the two types of complete unit cells can then be denoted by $f_{AB}(x) = f_A(x) + f_B(x)$ and $f_{BA}(x) = f_A(x - a/2) + f_B(x + a/2)$.

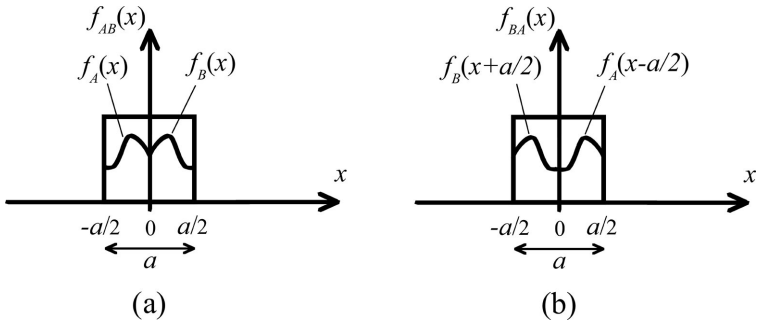

Figure 6.2 Two possible definitions of a unit cell of space group P1m in 1-D.

6.2.2 Diffraction by a Single Crystal

There are four varieties of 1-D crystals with N complete unit cells that can be constructed, as shown in Fig. 6.3, using the two molecules A and B. Their electron densities are labelled as $g^{(1)}(N, x)$ up to $g^{(4)}(N, x)$. The first two of these four crystal varieties, $g^{(1)}$ and $g^{(2)}$, are periodic whereas $g^{(3)}$ and $g^{(4)}$ are non-periodic because of the incomplete unit cells appended on their ends.

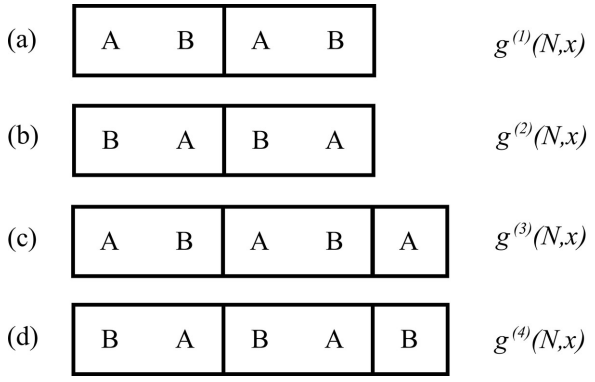

Figure 6.3 All possible types of crystals in 1-D with two molecules A and B in the unit cell. (a) and (b) have two full unit cells, (c) and (d) have two full and one incomplete unit cells.

The electron densities of the four varieties of crystals are

$$\begin{aligned}
g^{(1)}(N,x) &= f_{AB}(x) \otimes \sum_{h=0}^{N-1} \delta\left(x - ha + \frac{1}{2}(N-1)a\right) \\
g^{(2)}(N,x) &= f_{BA}(x) \otimes \sum_{h=0}^{N-1} \delta\left(x - ha + \frac{1}{2}(N-1)a\right) \\
g^{(3)}(N,x) &= f_{AB}(x) \otimes \sum_{h=0}^{N-1} \delta\left(x - ha + \frac{1}{2}(N-1)a\right) + f_A(x) \otimes \delta\left(x - \frac{1}{2}(N+1)a\right) \\
g^{(4)}(N,x) &= f_{BA}(x) \otimes \sum_{h=0}^{N-1} \delta\left(x - ha + \frac{1}{2}(N-1)a\right) + f_B(x) \otimes \delta\left(x - \frac{1}{2}(N+1)a\right).
\end{aligned} \tag{6.1}$$

The complex diffracted amplitudes in the far-field for the four crystal varieties are then

$$\begin{aligned}
G^{(1)}(N,u) &= F_{AB}(u)S(N,u) \\
G^{(2)}(N,u) &= F_{BA}(u)S(N,u) \\
G^{(3)}(N,u) &= F_{AB}(u)S(N,u) + F_A(u)\exp(i\pi(N+1)au) \\
G^{(4)}(N,u) &= F_{BA}(u)S(N,u) + F_B(u)\exp(i\pi(N+1)au),
\end{aligned} \tag{6.2}$$

where $S(N,u)$ is the shape transform and

$$F_{AB}(u) = F_A(u) + F_B(u) \tag{6.3}$$

$$F_{BA}(u) = F_A(u)\exp(i\pi au) + F_B(u)\exp(-i\pi au) \tag{6.4}$$

are the Fourier transforms of the two types of unit cell.

The corresponding diffracted amplitudes are

$$\begin{aligned}
I^{(1)}(N,u) &= |F_{AB}(u)|^2 S^2(N,u) \\
I^{(2)}(N,u) &= |F_{BA}(u)|^2 S^2(N,u) \\
I^{(3)}(N,u) &= |F_{AB}(u)|^2 S^2(N,u) + H^{(3)}(N,u)S(N,u) + |F_A(u)|^2 \\
I^{(4)}(N,u) &= |F_{BA}(u)|^2 S^2(N,u) + H^{(4)}(N,u)S(N,u) + |F_B(u)|^2,
\end{aligned} \tag{6.5}$$

where

$$\begin{aligned}
H^{(3)}(N,u) &= 2|F_A(u)|^2\cos(\pi(N+1)au) + 2Re\Big(F_A(u)F_B^*(u)\exp(i\pi(N+1)au)\Big) \\
H^{(4)}(N,u) &= 2|F_B(u)|^2\cos(\pi(N+2)au) + 2Re\Big(F_A(u)F_B^*(u)\exp(-i\pi Nau)\Big).
\end{aligned} \tag{6.6}$$

6.2.3 Ensemble-Averaged Diffraction

Assuming the crystal varieties $g^{(1)}$ and $g^{(2)}$ occur with equal probability, and similarly, $g^{(3)}$ and $g^{(4)}$ are equally likely to occur, a crystal size distribution $p(N)$ taking into account the presence of incomplete unit cells might resemble Fig. 6.4. Notice that instead of the probabilities occurring for only integer values of N as in the previous chapters, probabilities for half integers are now allowed due to the occurrence of incomplete unit cells.

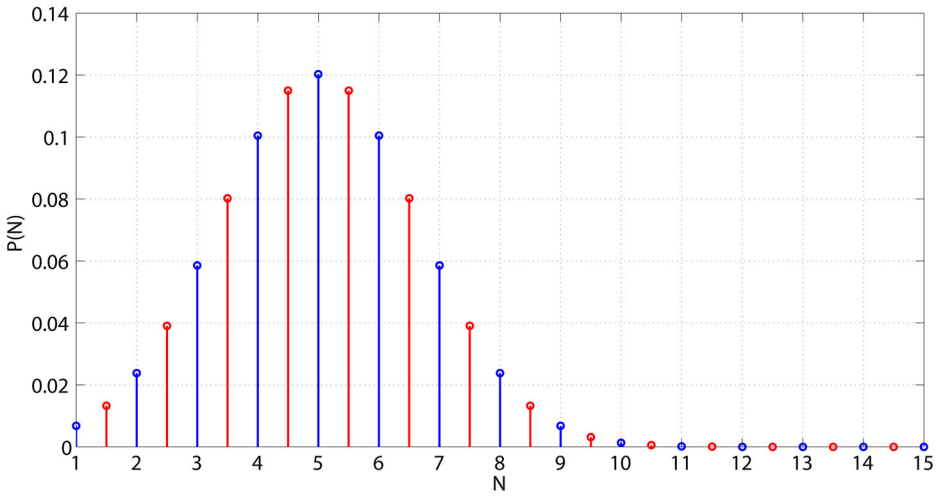

Figure 6.4 Truncated Gaussian crystal size distribution with a mean of 5 and standard deviation of 5/3 unit cells. Blue stems represent crystal sizes that are integers and red half-integers.

The averaged diffracted intensity over an ensemble of crystal sizes following a size distribution $p(N)$ is then given by

$$\begin{aligned} I(u) &= \langle I(N,u)\rangle_N \\ &= \frac{1}{2}\sum_N p(N)\left(I^{(1)}(N,u) + I^{(2)}(N,u)\right) \\ &\quad + \frac{1}{2}\sum_N p(N+1/2)\left(I^{(3)}(N,u) + I^{(4)}(N,u)\right). \end{aligned} \tag{6.7}$$

Note the index of the summations, N, is distinct from the crystallite size and is still an element of the integers hence the use of $p(N+1/2)$. Equation (6.7) can be readily expanded by substitution from Eq. (6.5). Grouping together the terms associated with the second, the first, and the zeroth power of the shape transform, and denoting them by I^B, I^{B2} and I^{B3}, respectively, the averaged diffracted intensity can be written as

$$I(u) = I^B(u) + I^{B2}(u) + I^{B3}(u), \tag{6.8}$$

where

$$
\begin{aligned}
I^B(u) &= \frac{1}{2}\left(|F_{AB}(u)|^2 + |F_{BA}(u)|^2\right)\sum_N \big(p(N)+p(N+1/2)\big)S^2(N,u) \\
I^{B2}(u) &= \frac{1}{2}\sum_N p(N+1/2)\left(H^{(3)}(N,u)+H^{(4)}(N,u)\right)S(N,u) \\
I^{B3}(u) &= \frac{1}{2}\left(|F_A(u)|^2 + |F_B(u)|^2\right)\sum_N p(N+1/2).
\end{aligned} \tag{6.9}
$$

The three additive components of the averaged diffracted intensity in Eq. (6.9) are referred to in this thesis as the Bragg (B), the "Bragg 2" ($B2$) and the "Bragg 3" ($B3$) components, respectively. The $B2$ component is not strictly an "intensity" for example it is not necessarily positive, but it represents the difference between the averaged diffracted intensity and the sum of the intensities diffracted by the full-unit-cell and incomplete unit-cell parts. The $B2$ component can therefore be thought of as the interference effect between full unit cells and incomplete unit cells situated on the edge of the crystal. By their definitions above, the Bragg component is of order N^2, the $B2$ component is order N, and the $B3$ component is order 1.

Inspection of the Bragg term in Eq. (6.9) shows that it can be written as

$$I^B(u) = \big\langle |F(u)|^2 \big\rangle Q^2(u) \tag{6.10}$$

where

$$\big\langle |F(u)|^2 \big\rangle = \frac{1}{2}\left(|F_{AB}(u)|^2 + |F_{BA}(u)|^2\right) \tag{6.11}$$

is the average molecular transform from the two kinds of unit cells AB and BA. The averaged squared shape transform in Eq. (6.10) is now given by

$$Q^2(u) = \sum_N (p(N)+p(N+1/2))S^2(N,u), \tag{6.12}$$

and the averaged diffracted intensity in Eq. (6.8) can be expressed as

$$I(u) = \big\langle |F(u)|^2 \big\rangle Q^2(u) + I^{B2}(u) + I^{B3}(u). \tag{6.13}$$

Moreover, because $\sum_N \left(p(N)+p(N+1/2)\right) = 1$ and $\sum_N p(N) \approx \sum_N p(N+1/2)$,

$$\sum_N p\big(N\big) \approx \sum_N p\big(N+\frac{1}{2}\big) \approx \frac{1}{2}, \tag{6.14}$$

and the $B3$ component in Eq. (6.9) can be approximated by

$$I^{B3}(u) \approx \frac{1}{4}\left(|F_A(u)|^2 + |F_B(u)|^2\right). \tag{6.15}$$

6.2.4 Effect on the Bragg Peaks

At the reciprocal lattice points, $u = u_h = h/a$ where $h \in \mathbb{Z}$. The averaged intensity is then $\langle I(N, u_h)\rangle_N = I^B(u_h) + I^{B2}(u_h) + I^{B3}(u_h)$, where

$$I^B(u_h) = |F_A(u_h) + F_B(u_h)|^2 \sum_N \left(p(N) + p(N+1/2)\right) N^2 \tag{6.16}$$

$$I^{B2}(u_h) = \Big(|F_A(u_h)|^2 + (-1)^h |F_B(u_h)|^2 + (1+(-1)^h) Re(F_A(u_h) F_B^*(u_h))\Big) \times \sum_N p(N+1/2)(-1)^{Nh} N \tag{6.17}$$

$$I^{B3}(u_h) = \frac{1}{2}\left(|F_A(u_h)|^2 + |F_B(u_h)|^2\right) \sum_N p(N+1/2). \tag{6.18}$$

For h even,

$$\begin{aligned} I^{B2}(u_h) &= \left(|F_A|^2 + |F_B|^2 + 2Re(F_A(u_h) F_B(u_h)^*)\right) \sum_N p(N+1/2)N. \\ &= |F_A(u_h) + F_B(u_h)|^2 \sum_N p(N+1/2)N, \end{aligned} \tag{6.19}$$

which has the same u_h-dependence as $I^B(u_h)$. For h odd,

$$I^{B2}(u_h) = \left(|F_A(u_h)|^2 - |F_B(u_h)|^2\right) \sum_N p(N+1/2)(-1)^N N. \tag{6.20}$$

This suggests that at the reciprocal lattice points for h odd, the $B2$ component is much smaller because of the term $\left(|F_A(u_h)|^2 - |F_B(u_h)|^2\right)$ and the $(-1)^N$ in the summation in Eq. (6.20). So it is conceivable that every second Bragg peak would be less affected by the effect of incomplete unit cells and using only those peaks in the average would produce a smaller error in the final averaged intensity compared to averaging over the entire set of measured Bragg reflections.

6.2.5 Effect on the Diffraction at General u

One-dimensional simulations are conducted to investigate the effect of incomplete unit cells on the ensemble-averaged diffraction at general reciprocal lattice locations. Employing the mirror symmetry relation of the space group P1m so that $f_A(x) = f_B(-x)$ and $F_A(u) = F_B^*(u)$, the molecular transforms for unit cell configurations AB and BA becomes

$$|F_{AB}(u)|^2 = 4Re^2(F_A(u)) \tag{6.21}$$

$$|F_{BA}(u)|^2 = 4\big(Re(F_A(u))\cos(\pi a u) - Im(F_A(u))\sin(\pi a u)\big)^2, \tag{6.22}$$

respectively.

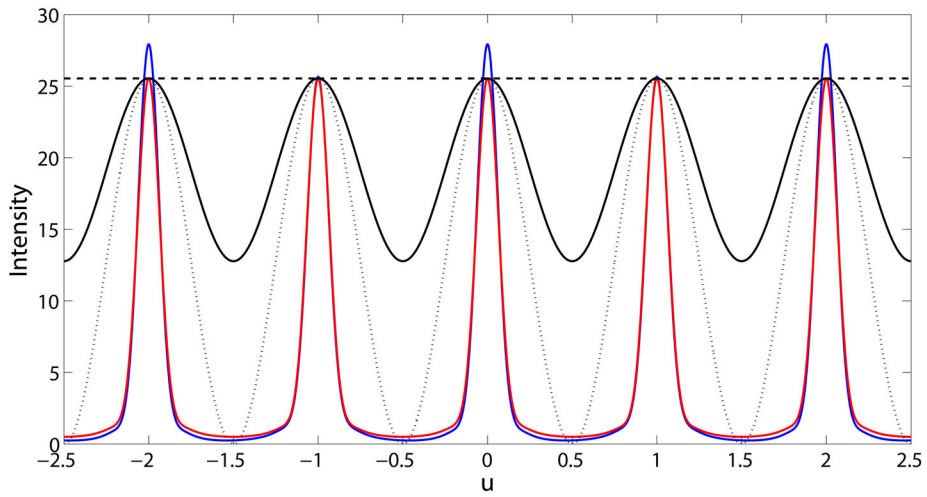

Figure 6.5 Averaged diffracted intensity $I = I^B + I^{B2} + I^{B3}$ (blue) for the 1-D electron density described in Sec. 6.2.5. The red curve shows the averaged diffracted intensity when no incomplete unit cells are present. The envelope of the intensities, i.e. the molecular transforms, $|F_{AB}|^2$ (dashed), $|F_{BA}|^2$ (dotted), and their average $\langle|F|^2\rangle$ (solid) are scaled to the red curve.

For the first simulation, the molecular transform for configuration AB is set to a constant to allow the effect of incomplete unit cells to be seen more clearly. Specifically, $F_A(u) = 1/2$ which implies $|F_{AB}(u)|^2 = 1$ and $|F_{BA}(u)|^2 = \cos^2(\pi a u)$ by Eqs. (6.21) and (6.22). The diffracted intensities $I^{(1)}$ to $I^{(4)}$ for this molecular transform are calculated via Eqs. (6.5) and shown in Fig. 6.6(a) for a single crystal of size $N = 5$. The difference between the intensity profiles around the even and odd Bragg reflections, i.e, when the value of h is even and odd respectively, are shown in Fig. 6.6(b) and (c).

For a Gaussian crystal size distribution with a mean of 5 and standard deviation 5/3 unit cells, as shown in Fig. 6.4, the averaged diffracted intensity over a collection of diffraction patterns from a variety of different-sized crystals with incomplete unit cells is calculated using Eq. (6.8) and shown in Fig. 6.5. A comparison is made with the averaged diffraction from a crystal ensemble without incomplete unit cells (red curve in Fig. 6.5) calculated by setting the $B2$ and $B3$ components to zero. It can be seen that the diffraction at even reciprocal lattice points are indeed less affected by the presence of incomplete unit cells than at odd reciprocal lattice points as predicted in the previous section. The B, $B2$ and $B3$ components are individually shown in Fig. 6.7.

The normalised summed error quantifying the discrepancy between the diffracted intensity at the Bragg peaks for an ensemble of crystals with incomplete unit cells, and an ensemble of crystals without incomplete unit cells is shown in Fig. 6.8. The decrease in error with increasing mean crystal size indicates that the effects of incomplete unit cells are smaller for larger crystals.

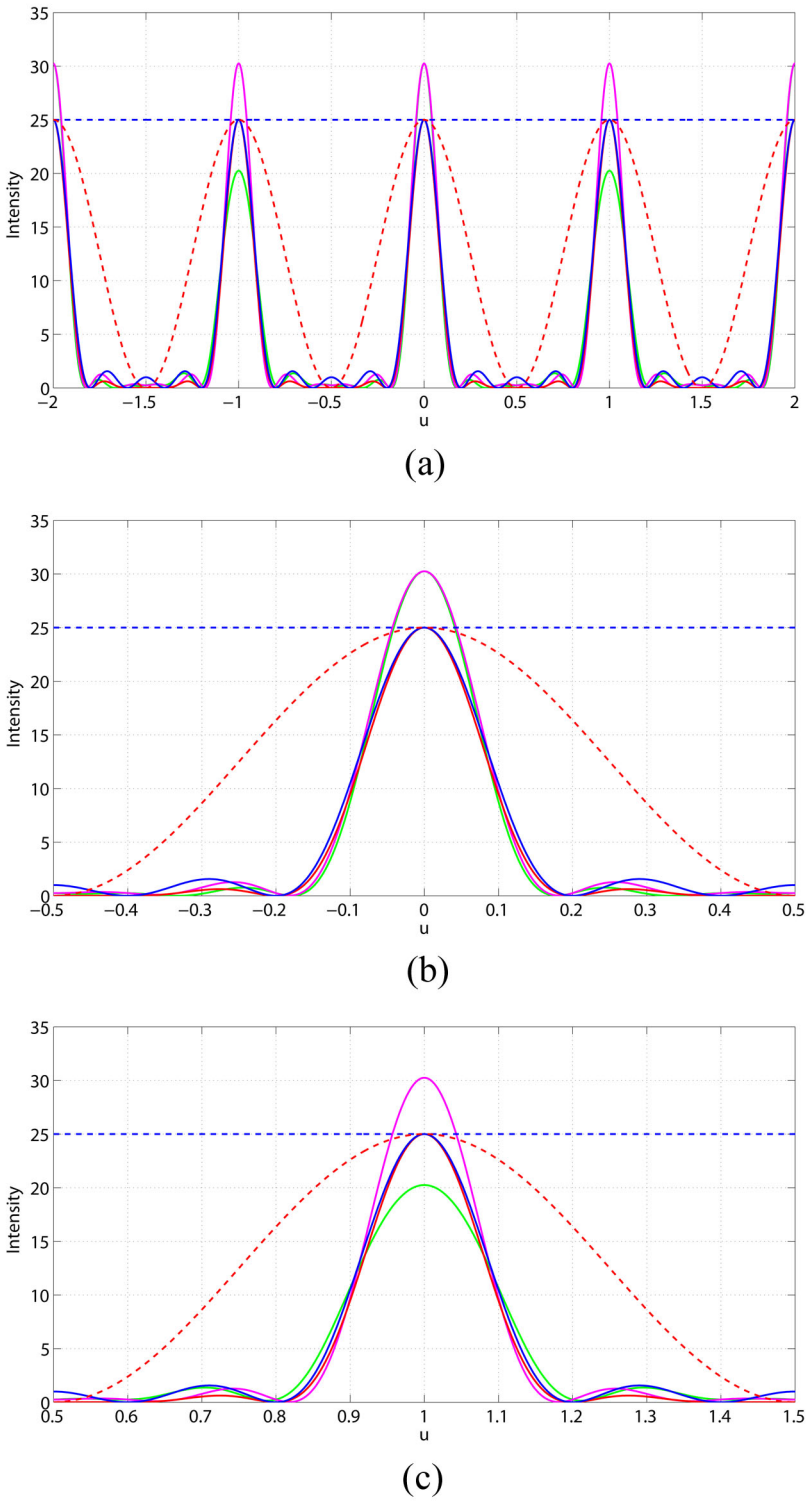

Figure 6.6 (a) Single-crystal diffracted intensities for the 1-D electron density described in Sec. 6.2.5 with $N = 5$ unit cells for the four different crystal variations, $I^{(1)}$ (blue), $I^{(2)}$ (red), $I^{(3)}$ (magenta) and $I^{(4)}$ (green). Envelopes (dashed curves) are the molecular transforms $|F_{AB}|^2$ (blue) and $|F_{BA}|^2$ (red). Enlarged view of Bragg reflection at (b) even and (c) odd reciprocal lattice points.

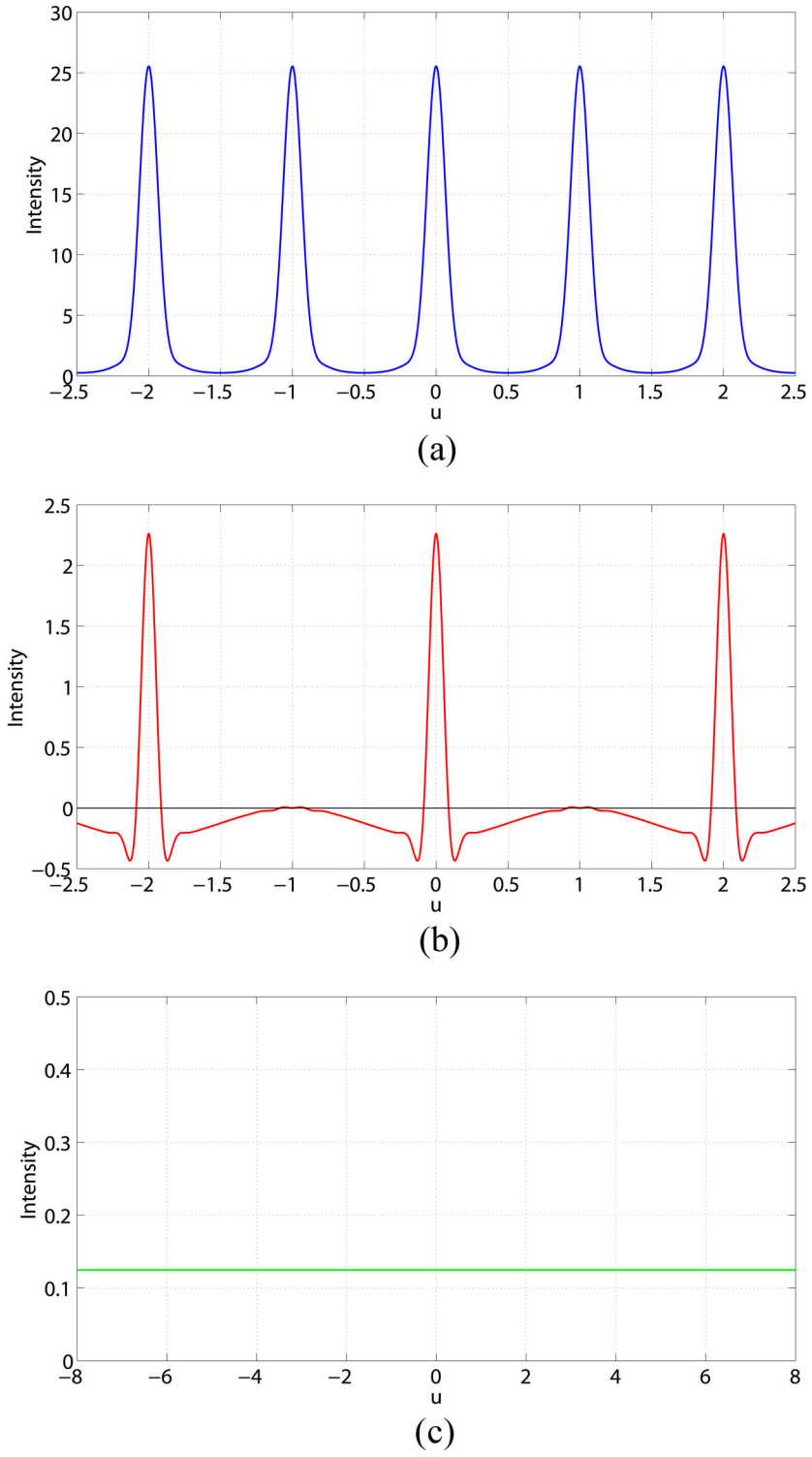

Figure 6.7 Components (a) I^B, (b) I^{B2} and (c) I^{B3} of the averaged diffracted intensity shown in Fig. 6.5.

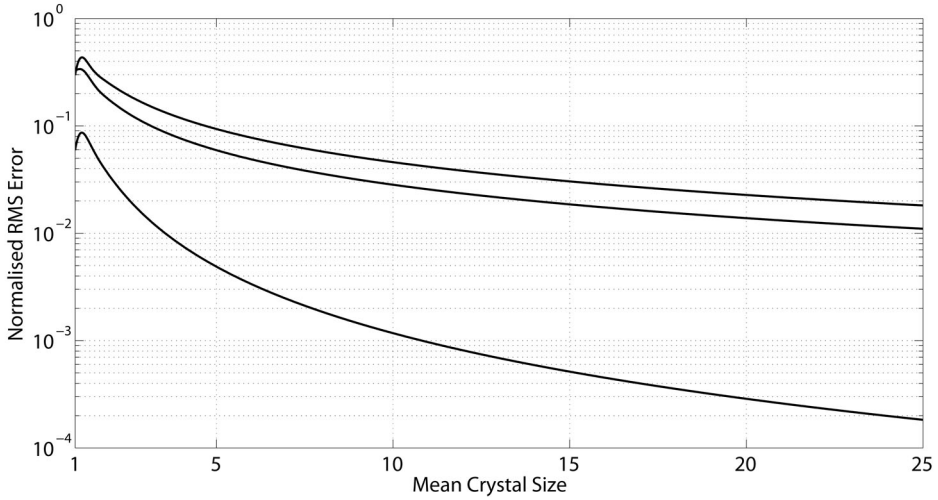

Figure 6.8 Root-mean-square error between the ensemble-averaged diffracted intensity over the Bragg peaks for a collection of crystals with incomplete unit cells and a collection of crystals without, and its dependence on mean crystallite size. Errors calculated with the set of Bragg peaks at even (top curve), odd (bottom curve) and all (middle curve) reciprocal lattice points.

6.2.6 Effect on Estimating the Molecular Transform

If the $B2$ and $B3$ components are small enough, then by Eqs. (6.8) and (6.10), the ensemble-averaged diffraction is approximately

$$I(u) \approx \big\langle |F(u)|^2 \big\rangle Q^2(u) \tag{6.23}$$

and $Q^2(u)$ can be estimated as described in Chapter 5 and $\langle |F(u)|^2 \rangle$ obtained by dividing $I(u)$ through by this estimate. The important result is that this calculation gives access only to $\langle |F(u)|^2 \rangle$, i.e. the diffracted intensity averaged over all possible arrangements of molecules in the unit cell. It is therefore not necessarily possible to extract the molecular transform intensity of a single unit cell or a single molecule. This is a fundamental limitation that, interestingly, is due not to the presence of incomplete unit cells, but to the different possible arrangements of molecules in the full unit cell.

A one-dimensional simulation is conducted to investigate these effects. A discrete signal with eight samples was constructed to model the 1-D electron density function of molecule A. Molecule B is then generated according to the relationship in space group P1m. The molecular transforms $|F_{AB}(u)|^2$, $|F_{BA}(u)|^2$ and their average $\langle |F(u)|^2 \rangle$, are calculated using the DFT and plotted in Fig. 6.9. The ensemble averaged diffraction for a collection of crystals with incomplete unit cells is again calculated via Eq. (6.8) for a Gaussian crystal size distribution with a mean of 5 and standard deviation of 5/3 unit cells. The total diffracted intensity $I(u)$ is shown as the blue curve in Fig. 6.9 and its three components are shown in Fig. 6.11. Inspection of the figure shows the nature of the three components and their relative magnitudes are as predicted in Sec. 6.2.3.

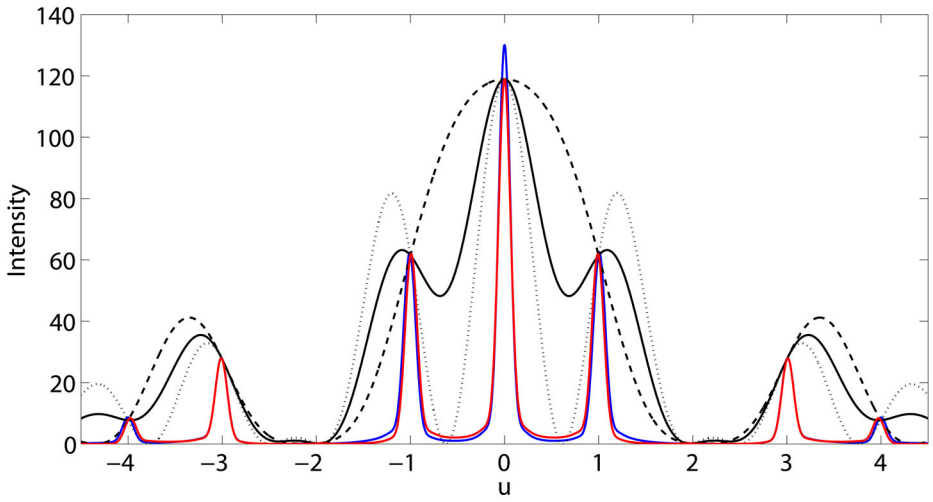

Figure 6.9 Averaged diffracted intensity (blue) for the 1-D electron density described in Sec. 6.2.6. The red curve shows the averaged diffracted intensity when no incomplete unit cells are present. The envelope of the intensities, i.e. the molecular transforms, $|F_{AB}|^2$ (dashed), $|F_{BA}|^2$ (dotted), and their average $\langle|F|^2\rangle$ (solid) are scaled to the red curve.

Estimation of the molecular transform is then simulated as follows. The Bragg volumes for the averaged diffracted intensity $I(u)$ in Fig. 6.9 are extracted and averaged together as described in Chapter 5. An estimate of the periodic component in $I(u)$, denoted by $\hat{Q}^2(u)$, is formed using the averaged Bragg volume. The quantity $\hat{Q}^2(u)$ equals the average squared shape transform when no incomplete unit cells are present. The average molecular transform is then estimated as

$$\langle|\hat{F}(u)|^2\rangle = \frac{I(u)}{\hat{Q}^2(u)}. \tag{6.24}$$

Applying this procedure to the averaged diffracted intensity in Fig. 6.9 gives an estimate of $\langle|\hat{F}(u)|^2\rangle$ as shown in Fig. 6.10. By the analysis from the previous chapter, the molecular transforms $|F_{AB}(u)|^2$ and $|F_{BA}(u)|^2$ individually tend towards a constant when summed over all Bragg volumes, and hence their average $\langle|\hat{F}(u)|^2\rangle$ will also approach a constant after summing enough Bragg volumes. Fourteen Bragg volumes were summed in total in this case. Inspection of Fig. 6.10 shows that a reasonably good estimate of the averaged molecular transform $\langle|F(u)|^2\rangle$ is obtained. A comparison is made with the estimated average molecular transform from a crystal ensemble without incomplete unit cells (red curve in Fig. 6.10). The slight discrepancy between the true averaged molecular transform and the estimation from a crystal ensemble without incomplete unit cells (the black and red curves, respectively) is due to the limited number of Bragg volumes summed and hence the averaged molecular transform being not yet constant over the resulting averaged Bragg volume.

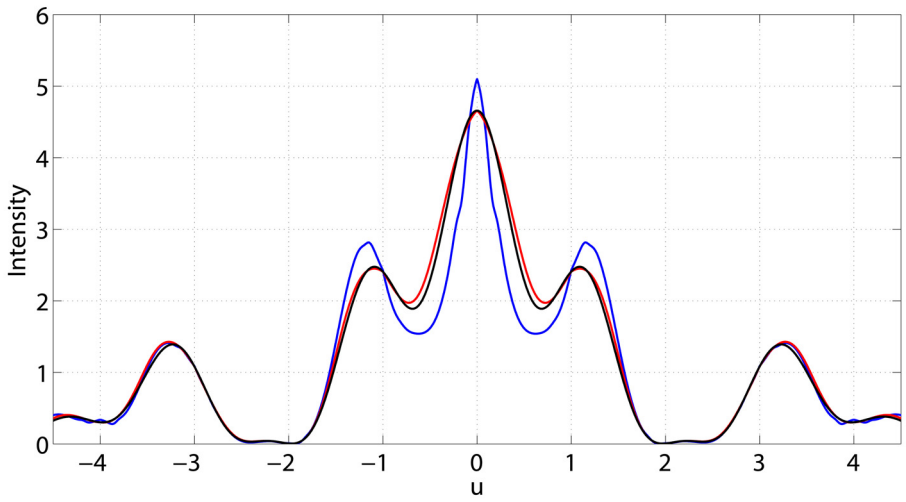

Figure 6.10 The recovered average molecular transform for a crystal ensemble with (blue) and without (red) incomplete unit cells. The true average molecular transform is shown in black.

6.2.7 Conclusions

The presence of multiple molecules in the unit cell and the incomplete unit cells that result affect the diffraction patterns of nanocrystals. The diffraction consists of a Bragg component, a "$B2$" component that is small but peaks at the reciprocal lattice points, and an even smaller "$B3$" component. The two important outcomes are that (1) the non-Bragg components are weak, even for small nanocrystals, and (2) only the molecular transform averaged over all molecular orientations in the unit cell can be estimated. Since the molecular transform itself cannot be estimated, this has significant implications for direct phasing if there are more than one molecule in the unit cell.

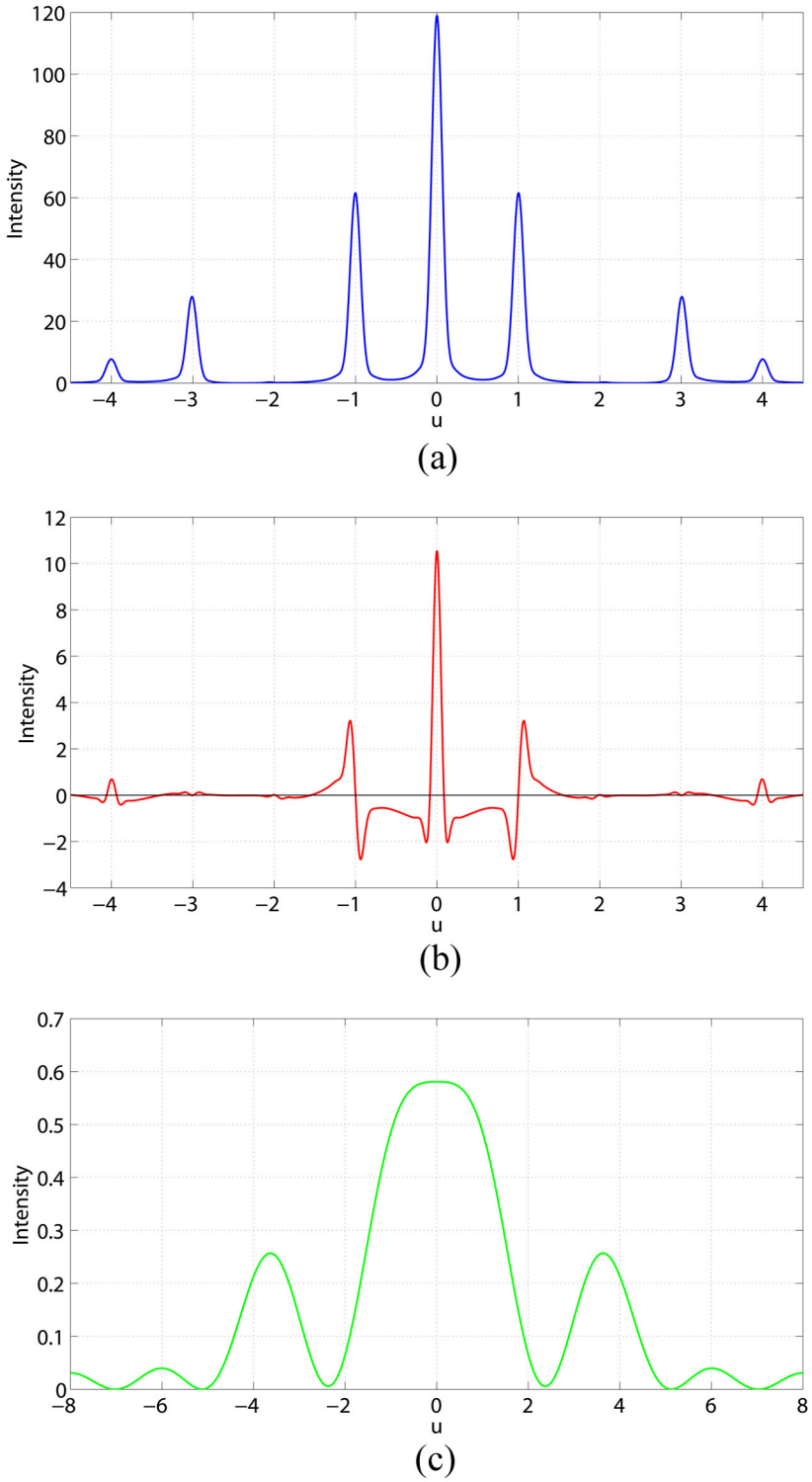

Figure 6.11 Components (a) I^B, (b) I^{B2} and (c) I^{B3} of the averaged diffracted intensity shown in Fig. 6.9.

6.3 NANOCRYSTALS IN TWO OR MORE DIMENSIONS

The richer topology of space in two or more dimensions means that the issues discussed in the previous section are much more complicated in higher dimensions. Furthermore, the many more symmetries and in turn, space groups that are allowed in 2-D and 3-D make the situation more complicated still. The case in 2-D is studied in detail in the following sections. The 3-D case will be similar although more intricate in detail.

6.3.1 Unit Cell Definition

Consider the case of a molecule that crystallises with more than one copy of itself in the unit cell. The positions and orientations of the molecule in the unit cell will generally be related by space-group symmetry. An infinite crystal is formed when repetitions of the unit cell are translated throughout a lattice. Recall that the electron density of a crystal, $g(\mathbf{x})$, can be written in terms of the electron density of the unit cell, $f(\mathbf{x})$, and a lattice matrix L of the underlying lattice, as

$$g(\mathbf{x}) = \sum_{\mathbf{m}} f(\mathbf{x} - \mathrm{L}\mathbf{m}), \tag{6.25}$$

where $\mathbf{m}$ is a vector of integers that indexes a particular position in the lattice.

If there is more than one molecule in the unit cell, then different unit cells, $f(\mathbf{x})$, can be defined. A natural redefinition of the unit cell is to modify its boundary such that the unit cell encompasses different spatial combinations of adjacent molecules in the crystal. As shown in the top row of Fig. 6.12 for a 2-D crystal with two molecules in the unit cell, this type of redefinition corresponds in the simplest case to shifting the boundary of the unit cell so that the contents of the different unit cells are translationally wrapped versions of each other. Crystals 1 and 2 in Fig. 6.12 then shows two crystals that can be described entirely with one type of unit cell while leaving incomplete unit cells when described by the other. Crystal 3 in the same figure illustrates the situation in practice where no one single kind of unit cell can be used to describe the crystal entirely, and incomplete unit cells inevitably remain. Upon enumeration of the different unit cells, the electron density of the crystal built from the jth kind of unit cell can be expressed as

$$g(\mathbf{x}) = \sum_{\mathbf{m}} f_j(\mathbf{x} - \mathrm{L}\mathbf{m}) + p_j(\mathbf{x}), \tag{6.26}$$

where $j = 1, \cdots, J$ indexes the different kinds of complete unit cell, and $p_j(\mathbf{x})$ is the incomplete unit cell part that arises on the surface of the crystal due to the redefinition of the unit cell. The different complete unit cells leave correspondingly different configurations of molecules on the surface of the crystal, giving rise to different configurations of incomplete unit cells $p_j(\mathbf{x})$, but are all equally valid to be used to construct the same crystal $g(\mathbf{x})$.

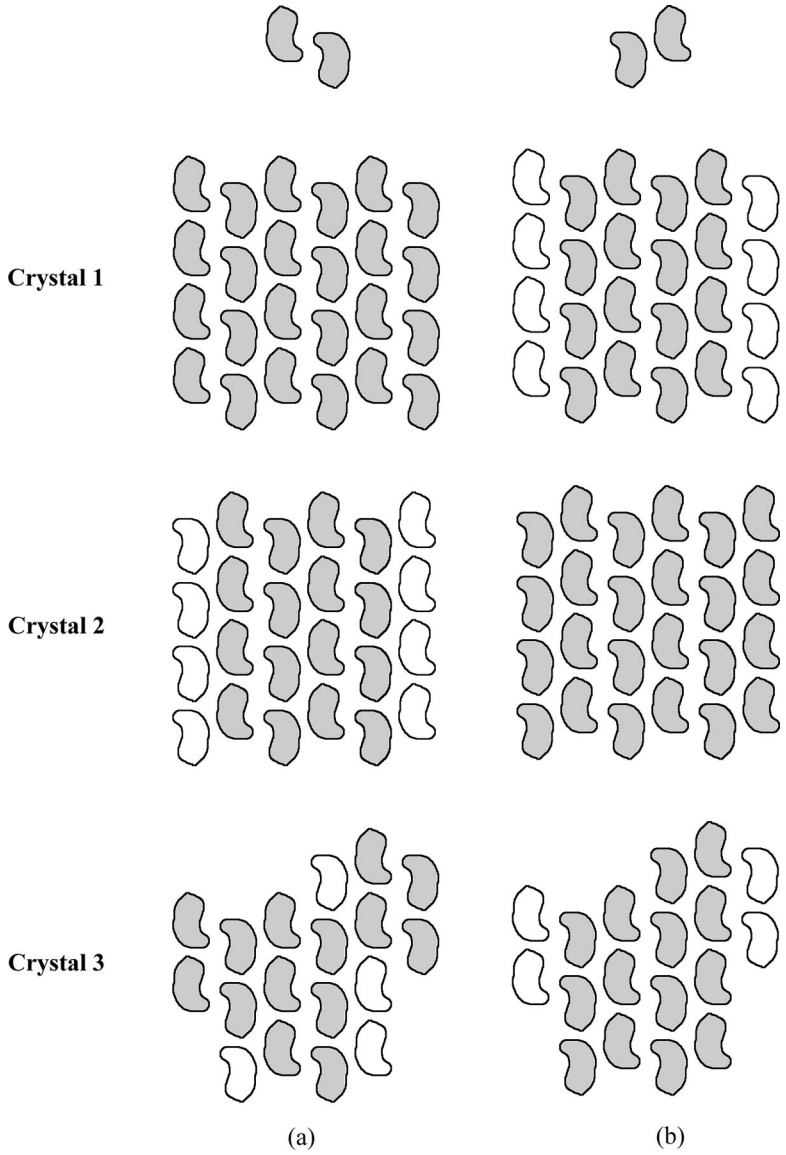

Figure 6.12 Two types of unit cells (top row) and three different crystals composed of the same molecular building blocks but terminated in different fashions (bottom three rows). (a) Description of the crystals using one kind of unit cell, (b) using another kind of unit cell - a translationally wrapped version of the unit cell from (a). Grey regions represent full unit cells and incomplete unit cells are indicated in white.

Note that for an infinite crystal, the particular assignment of the unit cell does not matter. One way to think about this is that there are no edges in an infinite crystal for there to be incomplete unit cells.

Figures. 6.13(a) and (b) show the symmetry relationships and possible unit cells for the case when four copies of a molecule (visualised as the letter "P") are arranged such that they have P4 symmetry, specifically, with a four-fold rotational axis about the centre of the unit cell highlighted in red in Fig. 6.13(a). The shaded area where the letter-P molecule resides is the asymmetric unit. The emergence of different possible unit cells is apparent from studying Fig. 6.13(a). For example, if the boundary of the

red unit cell is shifted half a unit cell distance up, or to the left, or both, encompassing the next whole asymmetric unit, then different unit cells result (blue, orange and green cells in Fig. 6.13(b), respectively). Figure 6.13(c) then shows four different 2-by-2 crystals that can be built using the four kinds of unit cells.

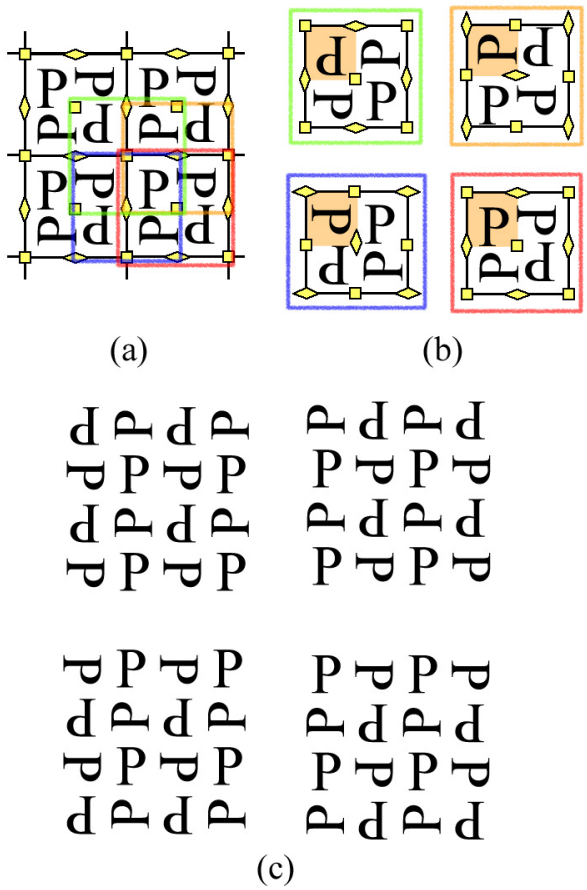

Figure 6.13 Different kinds of unit cells that can arise from shifting the boundary of a P4 crystal. (a) A 2-by-2 section of a 2-D crystal with associated space group symmetry symbols. (b) Four different kinds of unit cells that can be formed as described in the text. The shaded regions in orange represent the asymmetric unit. (c) 2-by-2 crystals built from the four different kinds of unit cells.

The different unit cells defined in this way have the interesting property that they do not have the same space group symmetry. For example, in Fig. 6.13(b), the red and green unit cells have P4 symmetry while the blue and orange cells are P2. This can be understood, and derived, by considering that the shifting of the unit cell boundary encompasses both the asymmetric units and the space group symbols, treating them as fixed in space with respect to the unit cell boundary shift. The change in space group symmetry for the same crystal as the unit cell is redefined is perplexing at first, however as we are no longer dealing with infinite, or even large crystals, the concept of space groups can no longer be applied - point group symmetries need to be considered instead.

A list of possible unit cells for distinct 2-D infinite crystal space groups determined by keeping the shape of the unit cell boundary the same as the one for the classical definition of that unit cell is shown in Fig. 6.14. For example, a 2-D nanocrystal containing two molecules labelled A and B arranged in P2 symmetry has two possible kinds of unit cell with the same shape as the original unit cell as shown in Fig. 6.14(b). Note that a multitude of other unit cells with different and sometimes more intricate shapes are also possible, but some of these may leave unphysical gaps or special facets on the edges of the crystal. More will be said on this in Sec. 6.4.

The space group symmetry of the corresponding infinite crystal and the relationships between the unit cells can be determined *a priori* from the diffraction data along with the crystal cell constants in practice. Averaging together diffraction patterns from many different crystals with different terminations in an SFX experiment thus generates a diffraction data-set that is averaged over some set of unit cells - a situation that is modelled next.

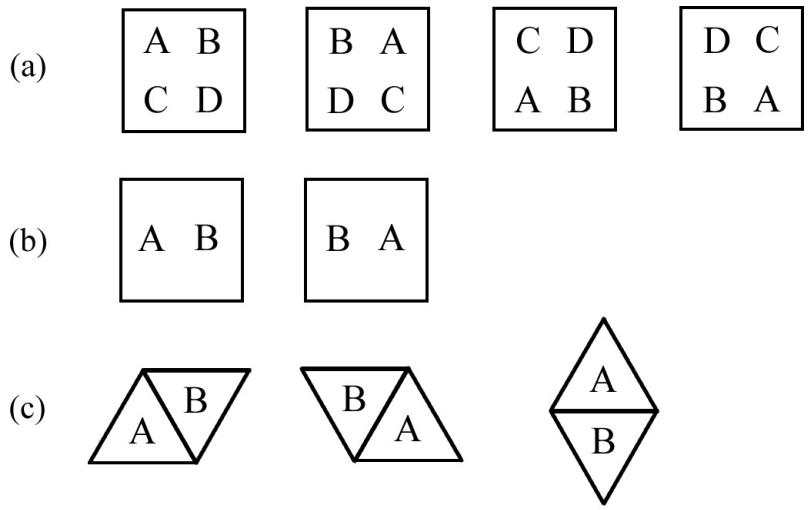

Figure 6.14 Different kinds of unit cells that can arise when the shape of the unit cell boundaries are kept the same as the one for the classical definition. (a) applies to P4, (b) to P2 and (c) to P3 and P6.

6.3.2 Modelling Random Nanocrystals

In general, regardless of the definition of the unit cell for a particular crystal, there will inevitably be a configuration of incomplete unit cells on the surface. A way to model the surface structure is required to describe the averaged diffraction from an ensemble of crystals with incomplete unit cells.

Consider the case of a 2-D crystal with P4 symmetry as described in the previous section. Assume that the full-unit-cell part of the crystal can be made up of one of a set of four different kinds of complete unit cell as illustrated in Fig. 6.14(a), that can all be used to build crystals with the same molecular arrangement.

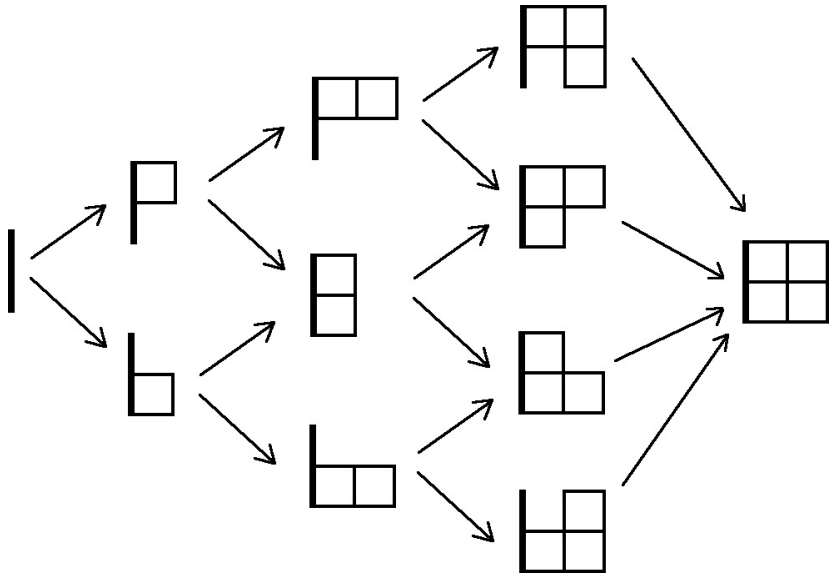

Figure 6.15 Constructive addition of PUCs. Each block represents a single asymmetric unit and the thick line is the edge of the part of the crystal composed entirely of complete unit cells.

In one approach [Chen and Millane 2013], the crystals are modelled from a constructive viewpoint in which each full-unit-cell crystal is first generated, and incomplete unit cells are then appended onto its surface in a random fashion. The process can be visualised in Fig. 6.15. This process generates random nanocrystals of the kind that is required. However, a shortcoming of this approach is that after the crystal has been generated, the complete unit cell part is not necessarily the largest that can be assigned to the crystal. Since the Bragg component of the diffraction is desired to be as large as possible in order to gauge the correct effect from incomplete unit cells, this is an important consideration.

Incomplete unit cells are always smaller than a complete unit cell, and only occur on the surface of the crystal. Thus, unless the crystal contains only one or two unit cells, it is dominated by the complete-unit-cell part, and in turn, the diffraction is dominated by the Bragg component that is due to this part. In general however, there is some flexibility in the separation of a crystal into the full-unit-cell and incomplete-unit-cell parts. The definition of the Bragg and Bragg-like components depends on this choice. Since the non-Bragg components, $B2$ and $B3$, can be considered as a perturbation on the Bragg component, it is more useful to construct the crystals in a way that the overall magnitude of the non-Bragg component is minimised. This can be achieved by minimising the total absolute value of the electron density in the incomplete unit cell part over the crystal ensemble. This can be done as follows. Referring to Fig. 6.16(a), addition of an incomplete unit cell that is larger than half a complete unit cell is equivalent to increasing the size of the full-unit-cell part of the crystal by one unit cell, and then subtracting an incomplete unit cell that is smaller than half of a complete unit cell. This mixture of addition and subtraction of incomplete unit cells using positive and

negative electron densities gives a smaller absolute value of the overall added electron density of the incomplete unit cells. Adding each incomplete unit cell in this way thus perturbs the full-unit-cell parts of the crystal less than the purely additive approach depicted in Fig. 6.16(b). The result is a larger, and often maximal, complete unit cell part.

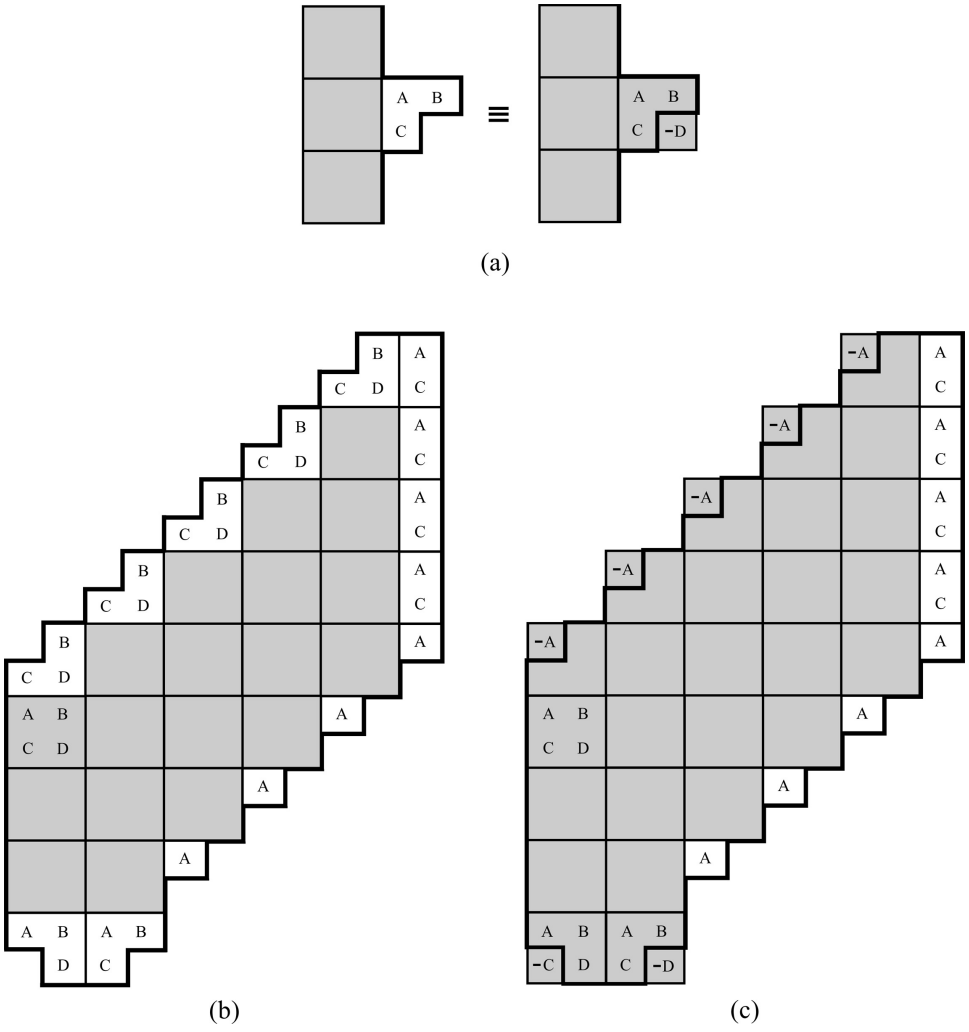

Figure 6.16 (a) Adding an incomplete unit cell consisting of 3/4 of a complete unit cell is equivalent to subtracting a smaller incomplete unit cell which is 1/4 of a complete unit cell from a full-unit-cell crystal that contains one more complete unit cell. (b) An example of a nanocrystal constructed by the purely additive approach. (c) An example of a nanocrystal constructed by a mixture of additive and subtractive approaches. The nanocrystal is shown bounded by the thick black lines. The full-unit-cell part is shaded grey, the added incomplete unit cells are shaded white, and the subtracted incomplete unit cells are indicated by a minus sign.

The model of the nanocrystal ensemble is therefore implemented in this way, and incomplete unit cells of maximum size half the complete unit cell, are added to or subtracted from a full-unit-cell crystal. An example of such a nanocrystal is illustrated in Fig. 6.16(c). The implications of this model for the ensemble-averaged diffraction are now investigated in detail.

6.3.3 An Ensemble of Crystals

Consider now a theoretical description for an ensemble of nanocrystals of different sizes, shapes and surface structures that applies in any dimensions. The ensemble is enumerated with indices (j,k,l). The index j indexes the kind of complete unit cell as described in Sec. 6.3.1 that makes up the full-unit-cell part of the crystal. The index k indexes the size and shape of the full-unit-cell part of the crystals. The index l indexes the surface structure, i.e., the arrangement of incomplete unit cells that are either added or subtracted on the surface of the full-unit-cell part of the crystal, which is referred to here as a *decoration*. Each nanocrystal is then indexed by the triple (j,k,l). Each decoration is composed of a total of R kinds of incomplete unit cell indexed by r, which is equivalent to the total number of ways a unit cell can be subdivided into its asymmetric units. Each kind of nanocrystal will occur with a different likelihood, which will be unknown.

The electron density of the (j,k,l) nanocrystal, $g_{jkl}(\mathbf{x})$, can now be written as

$$\begin{aligned} g_{jkl}(\mathbf{x}) &= f_j(\mathbf{x}) \otimes s_k(\mathbf{x}) + p'_{jkl}(\mathbf{x}) \\ &= f_j(\mathbf{x}) \otimes s_k(\mathbf{x}) + \sum_r p_r(\mathbf{x}) \otimes t_{klr}(\mathbf{x}) \ , \end{aligned} \tag{6.27}$$

where $f_j(\mathbf{x})$ is the electron density of the jth kind of complete unit cell, $s_k(\mathbf{x})$ is the shape function for the full-unit-cell part of the crystal that describes its underlying lattice, $p'_{jkl}(\mathbf{x})$ is the electron density of the (j,k,l) decoration, and $p_r(\mathbf{x})$ is the electron density of the rth type of incomplete unit cell. The positions of the rth type of incomplete unit cells, either added or subtracted, in the decoration are given by the lattice function $t_{klr}(\mathbf{x})$, which is referred to here as the *surface lattice*. Note that $t_{klr}(\mathbf{x})$ is independent of j because a particular arrangement of incomplete unit cells of type r in a particular crystal k fixes the kind of complete unit cell in the crystal, so that indexing by j is unnecessary.

The method of incomplete unit cell addition described in Sec. 6.3.2 leads to the notion of an outer surface lattice and an inner surface lattice as explained in the following. For a particular full-unit-cell crystal k, define an outer surface lattice, denoted $t_k^{(o)}(\mathbf{x})$, that consists of all lattice points, each with weight 1, on a single layer on the outside of the surface of the crystal (Fig. 6.17(a)). Define also an inner surface lattice, denoted $t_k^{(i)}(\mathbf{x})$, that consists of all lattice points, each with weight -1, on a single layer on the inside surface of the crystal (Fig. 6.17(a)). The two lattices and their associated $+1$ and -1 weights describe the addition and subtraction of the incomplete unit cells, respectively. The full surface lattice $t_k(\mathbf{x})$ can be written as the sum of these two lattices,

$$t_k(\mathbf{x}) = t_k^{(o)}(\mathbf{x}) + t_k^{(i)}(\mathbf{x}). \tag{6.28}$$

Note that these surface lattices are independent of j. The sites of the surface lattice

$t_{klr}(\mathbf{x})$ for the lth decoration and the rth kind of incomplete unit cell on the (j,k) full-unit-cell crystal are a subset of the sites of $t_k(\mathbf{x})$ such that for any l, there is one site from either $t_k^{(o)}(\mathbf{x})$ or $t_k^{(i)}(\mathbf{x})$, but not both, adjacent to a site on the (j,k) complete-unit-cell crystal surface, and that site is present for only one value of r. If each kind of incomplete unit cell is uniformly randomly either added to or subtracted from the full-unit-cell crystal, over a sufficiently large number of decorations l, there are on average an equal number of sites from $t_k^{(o)}(\mathbf{x})$ as from $t_k^{(i)}(\mathbf{x})$ in $t_{klr}(\mathbf{x})$ for any r.

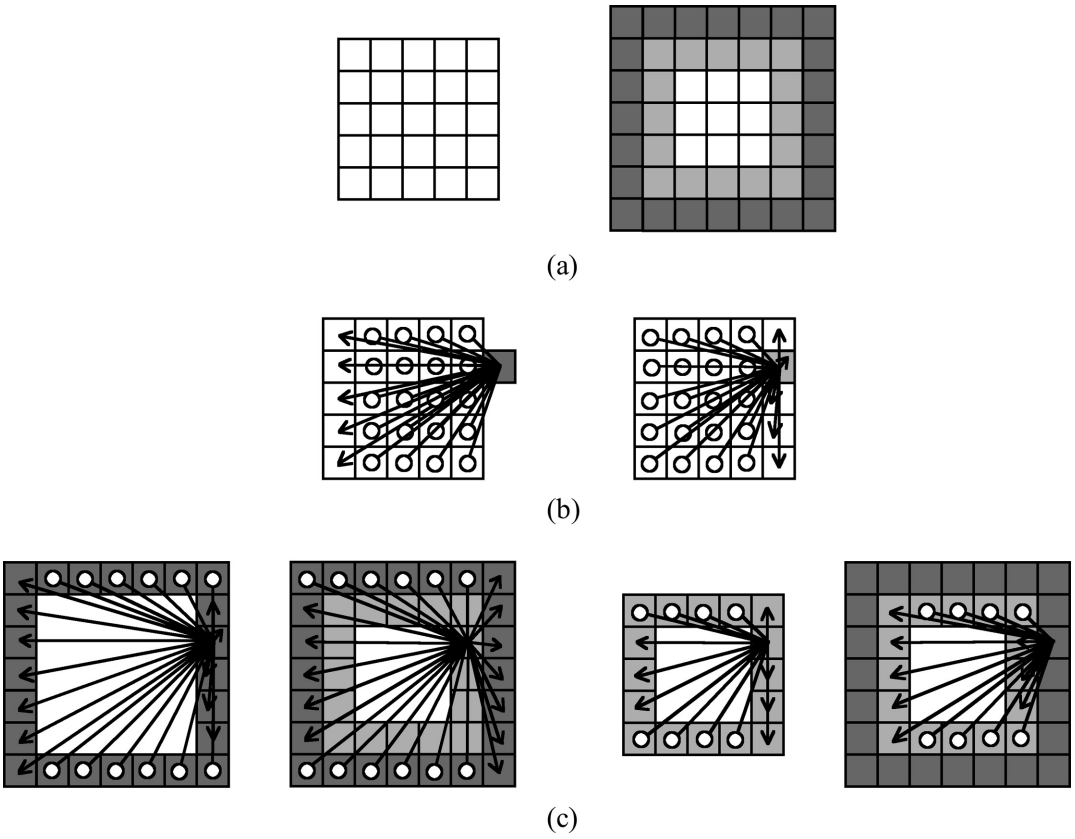

Figure 6.17 (a) A 5×5 full-unit-cell crystal (left) described by $s_k(\mathbf{x})$, and the corresponding outer (dark grey) and inner (light grey) full surface lattices $t_k^{(o)}(\mathbf{x})$ and $t_k^{(i)}(\mathbf{x})$, respectively (right). (b) Intersite vectors for a single site between $s_k(\mathbf{x})$ and $t_k^{(o)}(\mathbf{x})$ (left), and $s_k(\mathbf{x})$ and $t_k^{(i)}(\mathbf{x})$ (right) as described in the text. (c) Intersite vectors for a single site within $t_k^{(o)}(\mathbf{x})$ (left), $t_k^{(i)}(\mathbf{x})$ (middle-right), and from $t_k^{(i)}(\mathbf{x})$ to $t_k^{(o)}(\mathbf{x})$ (middle-left) and $t_k^{(o)}(\mathbf{x})$ to $t_k^{(i)}(\mathbf{x})$ (right). Intersite vectors that occur in both cases for (b), so that the corresponding terms in Eq. (6.32) cancel, and similarly in the left and right cases for (c) and the terms in Eq. (6.33) cancel, are shown by the line segments with empty circles. All intersite vector that occur in one case but not the other, and so the corresponding terms in Eq. (6.32) and (6.33) do not cancel, are shown by the arrows.

The diffracted intensity averaged over the ensemble of nanocrystals, $I(\mathbf{u}) = \langle |G_{jkl}(\mathbf{u})|^2 \rangle_{jkl}$, can now be written as

$$\begin{aligned} I(\mathbf{u}) &= I^B(\mathbf{u}) + I^{B2}(\mathbf{u}) + I^{B3}(\mathbf{u}) \\ &= \langle |F_j(\mathbf{u})|^2 |S_k(\mathbf{u})|^2 \rangle_{jk} \\ &\quad + \left\langle 2Re\Big(F_j(\mathbf{u}) S_k(\mathbf{u}) \sum_r P_r^*(\mathbf{u}) T_{klr}^*(\mathbf{u}) \Big) \right\rangle_{jkl} \\ &\quad + \left\langle \sum_r \sum_{r'} P_r(\mathbf{u}) P_{r'}^*(\mathbf{u}) T_{klr}(\mathbf{u}) T_{klr'}^*(\mathbf{u}) \right\rangle_{kl} , \end{aligned} \tag{6.29}$$

where the upper case quantities denote the Fourier transform of the corresponding lower case quantities. Equation (6.29) can be reduced to

$$\begin{aligned} I(\mathbf{u}) &= \langle |F_j(\mathbf{u})|^2 \rangle_j \, \langle |S_k(\mathbf{u})|^2 \rangle_k + 2Re\big(\langle F_j(\mathbf{u}) P_r^*(\mathbf{u}) \rangle_{jr} \big) \, \langle S_k(\mathbf{u}) T_{klr}^*(\mathbf{u}) \rangle_{klr} \\ &\quad + \langle P_r(\mathbf{u}) P_{r'}^*(\mathbf{u}) \rangle_{rr'} \, \langle T_{klr}(\mathbf{u}) T_{klr'}^*(\mathbf{u}) \rangle_{klrr'} \\ &= \langle |F_j(\mathbf{u})|^2 \rangle_j \, Q_B^2(\mathbf{u}) + 2Re\big(\langle F_j(\mathbf{u}) P_r^*(\mathbf{u}) \rangle_{jr} \big) \, Q_{B2}^2(\mathbf{u}) \\ &\quad + \langle P_r(\mathbf{u}) P_{r'}^*(\mathbf{u}) \rangle_{rr'} \, Q_{B3}^2(\mathbf{u}) , \end{aligned} \tag{6.30}$$

so that

$$\begin{aligned} I^B(\mathbf{u}) &= \langle |F_j(\mathbf{u})|^2 \rangle_j \, Q_B^2(\mathbf{u}) \\ I^{B2}(\mathbf{u}) &= 2Re\big(\langle F_j(\mathbf{u}) P_r^*(\mathbf{u}) \rangle_{jr} \big) \, Q_{B2}^2(\mathbf{u}) \\ I^{B3}(\mathbf{u}) &= \langle P_r(\mathbf{u}) P_{r'}^*(\mathbf{u}) \rangle_{rr'} \, Q_{B3}^2(\mathbf{u}) , \end{aligned} \tag{6.31}$$

where $Q_B^2(\mathbf{u})$, $Q_{B2}^2(\mathbf{u})$ and $Q_{B3}^2(\mathbf{u})$ are the averaged shape transform functions for the three components. It is assumed that $s_k(\mathbf{x})$ and $t_{klr}(\mathbf{x})$, and their inversion symmetric copies, i.e. $s_k(-\mathbf{x})$ and $t_{lkr}(-\mathbf{x})$, occur equally frequently in the ensemble, so that the averaged shape transform functions are real.

The nature of each diffraction component is described by its averaged shape transform function. The averaged squared shape transform $Q_B^2(\mathbf{u})$ for the Bragg component is the usual averaged shape transform for a full-unit-cell crystal. Consider first the contribution to $Q_{B2}^2(\mathbf{u})$ for a single full-unit-cell crystal averaged over all decorations, i.e., the quantity $Q_{B2,k}^2(\mathbf{u}) = \langle S_k(\mathbf{u}) T_{klr}^*(\mathbf{u}) \rangle_{lr}$, which can be written as

$$Q_{B2,k}^2(\mathbf{u}) \propto \sum_{l,r,m,n} \Big(\exp(i2\pi\mathbf{u} \cdot (\mathbf{a}_{km} - \mathbf{b}_{klrn}^{(o)})) - \exp(i2\pi\mathbf{u} \cdot (\mathbf{a}_{km} - \mathbf{b}_{klrn}^{(i)})) \Big) , \tag{6.32}$$

where $\mathbf{a}_{km}$ denotes the positions of the sites of the full-unit-cell lattice, and $\mathbf{b}_{klrn}^{(o)}$ and $\mathbf{b}_{klrn}^{(i)}$ denote the positions of the sites of the surface lattice $t_{klr}(\mathbf{x})$ that belong to $t_k^{(o)}(\mathbf{x})$ and $t_k^{(i)}(\mathbf{x})$, respectively. This function depends on the vectors between the sites of the

full-unit-cell lattice and the full outer and inner surface lattices. Some of these two sets of intersite vectors are shown in Fig. 6.17(b), which shows, for simplicity, a square full-unit-cell crystal (the argument generalises to other shaped crystals on averaging over an ensemble of crystal shapes). Inspection of the figure shows that in most cases the same vectors occur in both sets and so the corresponding terms in Eq. (6.32) cancel. The intersite vectors that do not occur in both terms and so contribute to $Q^2_{B2,k}(\mathbf{u})$ are shown by the arrows in Fig. 6.17(b). For a crystal that has N lattice sites in the full-unit-cell part and M sites on the surface, there are nominally $2MN$ terms in Eq. (6.32), but reference to Fig. 6.17(b) shows that only $M^2/2$ in 2-D or $M^2/3$ in 3-D remain after cancellation. The important result is that the number of terms contributing to $Q^2_{B2,k}(\mathbf{u})$ is reduced from $O(MN)$ to $O(M^2)$ as a result of the cancellation. Since $Q^2_{B2}(\mathbf{u}) = \langle Q^2_{B2,k}(\mathbf{u})\rangle_k$, this carries over to the averaged shape transform function $Q^2_{B2}(\mathbf{u})$ with N and M being replaced by the average number of complete and incomplete unit cells in the ensemble, respectively. Note that for the sites of $t_{klr}(\mathbf{x})$ belonging to $s_k(\mathbf{x})$, i.e. those given by $\mathbf{b}^{(i)}_{klrn}$, they are coincident with one $\mathbf{a}_{km}$ and the argument of the second exponential in Eq. (6.32) vanishes. These terms contribute a negative constant for each (k,l,r), proportional to $-\bar{M}$, and so to $Q^2_{B2}(\mathbf{u})$. Note also that at any reciprocal lattice point, the arguments of all the exponentials in Eq. (6.32) vanish since all the terms cancel out, so that $Q^2_{B2}(\mathbf{u})$, and consequently $I^{B2}(\mathbf{u})$, vanishes at the Bragg reflections.

Consider now the averaged shape function $Q^2_{B3}(\mathbf{u})$ which can be written in a form similar to Eq. (6.32), given by

$$Q^2_{B3,k}(\mathbf{u}) \propto \sum_{l,r,r',n,n'} \Big(\exp(i2\pi\mathbf{u}\cdot(\mathbf{b}^{(o)}_{klrn} - \mathbf{b}^{(o)}_{klr'n'})) - \exp(i2\pi\mathbf{u}\cdot(\mathbf{b}^{(o)}_{klrn} - \mathbf{b}^{(i)}_{klr'n'})) - \exp(i2\pi\mathbf{u}\cdot(\mathbf{b}^{(i)}_{klrn} - \mathbf{b}^{(o)}_{klr'n'})) + \exp(i2\pi\mathbf{u}\cdot(\mathbf{b}^{(i)}_{klrn} - \mathbf{b}^{(i)}_{klr'n'})) \Big) . \quad (6.33)$$

The expression for $Q^2_{B3}(\mathbf{u})$ contains terms involving intersite vectors within and between sites of the outer and inner surface lattices. For a square crystal, there is a total of $4M^2$ such terms in 2-D and $2M^2$ in 3-D, and conducting a similar analysis to that for $Q^2_{B2}(\mathbf{u})$, shows that in this case most terms do not cancel, as shown in Fig. 6.17(c), and the number of terms contributing remains $O(M^2)$. The overall amplitude of the $B3$ component is therefore of the same order as the $B2$ component. At the reciprocal lattice points, for the same reasons as described above, many, but not all, of the corresponding terms cancel, and $I^{B3}(\mathbf{u})$ is small, but non-zero.

The average $B2$ and $B3$ intensity components relative to the B component is relevant and is related to the number of incomplete unit cells on the surface compared to those in the complete-unit-cell part of the crystals. These ratios can be estimated as follows. For a crystal with N unit cells, the diffracted intensity averaged over reciprocal space, $\langle I(\mathbf{u})\rangle_{\mathbf{u}}$, is proportional to the square of the integrated electron density of the

unit cell and to N (or to the square root of the number of terms in the interference function) as shown in Sec. 4.2.1. For the case of a 3-D cubic crystal with N unit cells, the number of surface lattice sites is $M = 6N^{2/3}$. The average integrated electron density of an incomplete unit cell in the model is 1/4 that of a complete unit cell, and the outer and inner surface lattices are each occupied with probability 1/2. Thus, utilising Eq. (6.31) and normalising the integrated electron density of the complete unit cell to 1, so that the integrated density of the incomplete unit cell is 1/4, gives

$$\begin{aligned}
\langle I^B(\mathbf{u})\rangle_{\mathbf{u}} &\approx \langle F(\mathbf{u})\rangle_{\mathbf{u}}\langle F(\mathbf{u})\rangle_{\mathbf{u}}\sqrt{N^2} = 1\cdot 1\ N = N \\
\langle I^{B2}(\mathbf{u})\rangle_{\mathbf{u}} &\approx \langle F(\mathbf{u})\rangle_{\mathbf{u}}\langle P(\mathbf{u})\rangle_{\mathbf{u}}\sqrt{\frac{M^2}{3}\cdot\frac{1}{2}} = 1\cdot\frac{1}{4}\,M\sqrt{\frac{1}{3}\cdot\frac{1}{2}} = \frac{\sqrt{6}}{4}N^{2/3} \\
\langle I^{B3}(\mathbf{u})\rangle_{\mathbf{u}} &\approx \langle P(\mathbf{u})\rangle_{\mathbf{u}}\langle P(\mathbf{u})\rangle_{\mathbf{u}}\sqrt{2M^2\cdot\frac{1}{2}} = \frac{1}{4}\cdot\frac{1}{4}\,M\sqrt{2\cdot\frac{1}{2}} = \frac{3}{8}N^{2/3}
\end{aligned} \tag{6.34}$$

for 3-D crystals. The relative average intensities of the $B2$ and $B3$ components to the B component for an ensemble of crystals of varying sizes are then

$$\begin{aligned}
\frac{\langle I^{B2}(\mathbf{u})\rangle_{\mathbf{u}}}{\langle I^{B}(\mathbf{u})\rangle_{\mathbf{u}}} &\approx \frac{\sqrt{6}}{4}\bar{N}^{-1/3} \\
\frac{\langle I^{B3}(\mathbf{u})\rangle_{\mathbf{u}}}{\langle I^{B}(\mathbf{u})\rangle_{\mathbf{u}}} &\approx \frac{3}{8}\bar{N}^{-1/3},
\end{aligned} \tag{6.35}$$

where N is replaced by $\bar{N}$, the mean number of unit cells in the ensemble of crystals. These ratios are plotted in Fig. 6.18. The $B2$ and $B3$ components have similar average amplitudes with $\langle I^{B2}(\mathbf{u})\rangle_{\mathbf{u}}/\langle I^{B3}(\mathbf{u})\rangle_{\mathbf{u}} \approx 2\sqrt{6}/3 \approx 1.6$. Note that these are averages over all reciprocal space and the precise ratio of the components may vary widely depending on $\mathbf{u}$. A similar analysis gives the following estimates for 2-D crystals

$$\begin{aligned}
\frac{\langle I^{B2}(\mathbf{u})\rangle_{\mathbf{u}}}{\langle I^{B}(\mathbf{u})\rangle_{\mathbf{u}}} &\approx \frac{1}{2}\bar{N}^{-1/2} \\
\frac{\langle I^{B3}(\mathbf{u})\rangle_{\mathbf{u}}}{\langle I^{B}(\mathbf{u})\rangle_{\mathbf{u}}} &\approx \frac{1}{4}\bar{N}^{-1/2}\,,
\end{aligned} \tag{6.36}$$

with $\langle I^{B2}(\mathbf{u})\rangle_{\mathbf{u}}/\langle I^{B3}(\mathbf{u})\rangle_{\mathbf{u}} \approx 2$.

6.3.4 Simulation

The theory described above is illustrated by numerical simulation of ensembles of 2-D nanocrystals. A square image with plane group symmetry P4 as shown in the top left of Fig. 6.19 is used as the unit cell. The three other kinds of complete unit cell that can be generated by translational wrapping, as described in Sec. 6.3.1, are shown in the same figure. Two types of full-unit-cell crystal shapes, rectangular and random, were generated as follows.

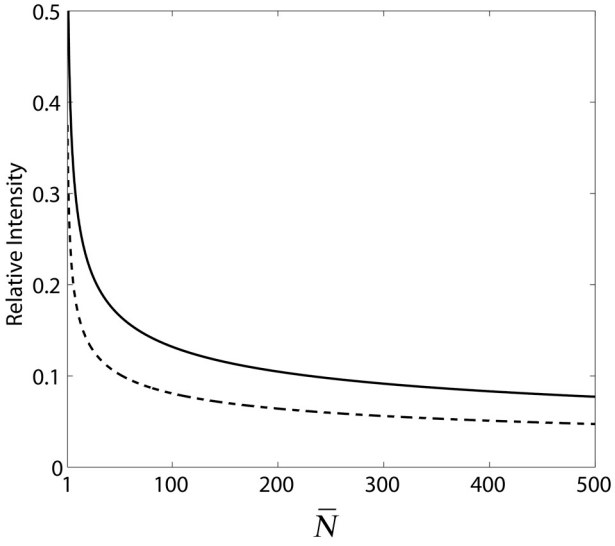

Figure 6.18 Average amplitudes of the $B2$ (solid line) and $B3$ (dashed line) intensities relative to the B component versus the mean number of unit cells in a 3-D crystal as described by Eq. (6.35).

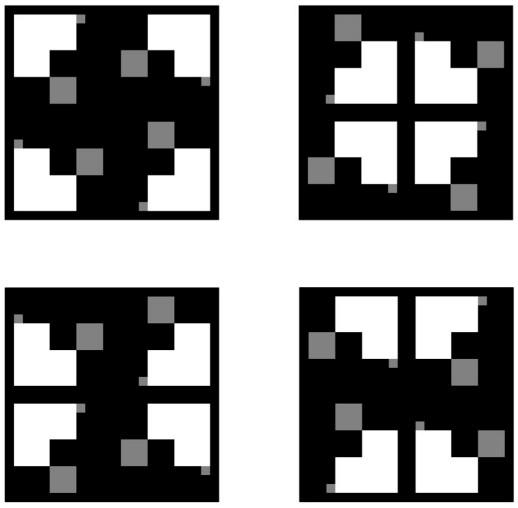

Figure 6.19 The unit cell used in the simulations showing the four corresponding kinds of full unit cell. The values represented are 0 (black), 1 (grey) and 2 (white).

6.3.4.1 Growing Crystals

For the generation of randomly shaped crystals, the total number of unit cells N, is first chosen from a distribution of integers, a sampled Gaussian in this case, truncated for $N < 1$ with a specified mean $\bar{N}$ and standard deviation σ_N. A square matrix of zeroes is then initialised as the "world" for the crystal to grow, with side lengths $2N+1$ to accommodate all possible crystal shapes. The zeroth iteration consists of placing one unit cell at the centre of the world matrix and each iteration thereafter adds an additional unit cell uniformly randomly to one of the sides of the current shape. This is done by tallying up the total number of sides that the current crystal shape possesses and assigning equal probabilities to each of those sides. All possible crystal shape combinations for $N = 1$ to 4 are shown in Fig. 6.20. This way of growth means that for a potential unit cell location that shares two, three, or four edges with the current crystal shape, a new cell will be 2, 3 or 4 times more likely, respectively, to be added there in the next iteration compared to a cell sharing only one edge with the current crystal shape. The next growth position is solely chosen by selecting an edge of the current crystal shape at random. Thus, as N becomes large, this crystal generating method will tend to produce spherical (circular in 2-D), clusters of crystals. The case of a potential cell position having four edges in common with the current shape indicates that a hole the size of a single cell site is present in the crystal. Single cell site holes are likely to be subsequently filled in by the crystal growth process described above but if survived after the growing step, they would contribute to the diffraction pattern in a manner akin to substitutional defects in real crystals. The process described has been studied by Eden [1961] in the context of modelling cellular growth in biology and is now known as the Eden model.

For the generation of rectangular crystals, the lengths of the two edges are each chosen from again a normal distribution of integers, truncated below unity, with a specified mean and standard deviation. No dynamic growth process is needed. The two different types of crystal shapes are desired as rectangular crystals show the effects of faceted crystal boundaries while the random-shaped crystals show the effects of rougher boundaries.

After the shape of the full-unit-cell crystal has been generated, incomplete unit cells are added to, or subtracted from, the surface of the crystal. For each surface position on the full-unit-cell crystal, a site is randomly selected from either the adjacent outer or inner surface lattice sites. An incomplete unit cell type is then randomly selected from the set of compatible incomplete unit cells (including the empty incomplete unit cell), and is incorporated at that site. By "compatible" it is meant that the incomplete unit cell is oriented correctly with respect to the crystal edge that it is to be associated with. Note that, as described in Sec. 6.3.2, the incomplete unit cells are no larger than half the size of the complete unit cells. A list of incomplete unit cell configurations that was

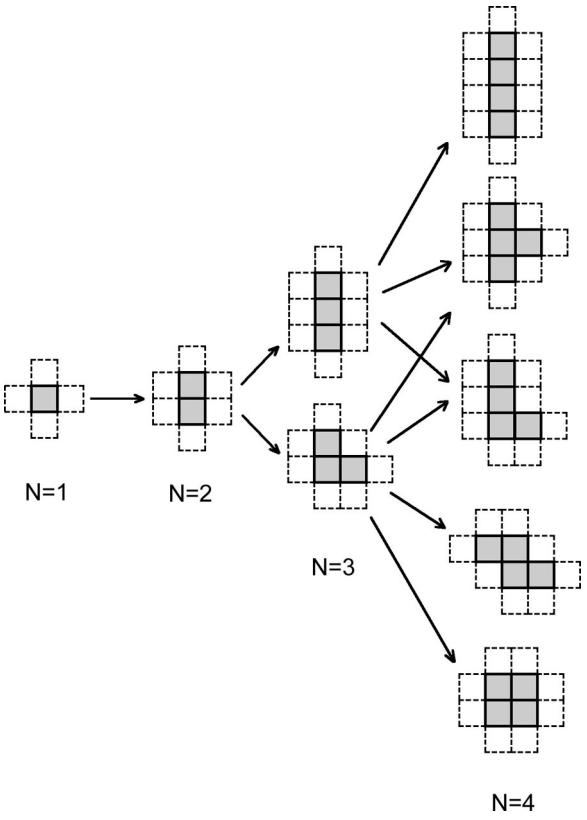

Figure 6.20 A random crystal growing process. The square unit cells multiply according to the rule that additional cells gets tacked on to the sides of the existing crystal (grey) with probabilities proportional to the number of existing cells the potential growth site (dotted) is adjacent to. The number of unique shapes that a crystal can have, called *polyominoes*, grows exponentially with the size of the crystal, given by the total number of unit cells N.

employed for the P4 case in this simulation is shown in Fig. 6.21. Other configurations are possible, as was shown in Fig. 6.15, but it was decided that the seven depicted in Fig. 6.21 were the most stable and hence the most probable to occur in practice. Examples of crystals obtained after the above processes are shown in Fig. 6.22.

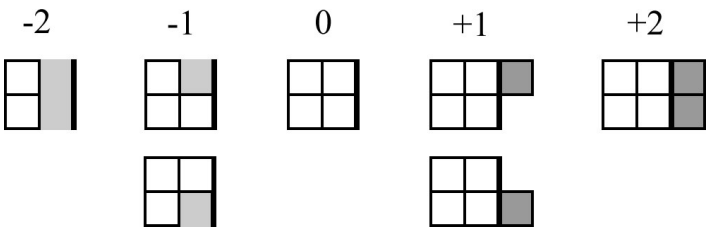

Figure 6.21 Seven possible configurations of incomplete unit cells in 2-D in one particular direction out of four possible directions. The thick line represents the edge of the crystal before the incorporation of incomplete unit cells. Regions subtracted are shown in light grey and regions added are in dark grey. The numbers represent the net number of asymmetric units added for each configuration.

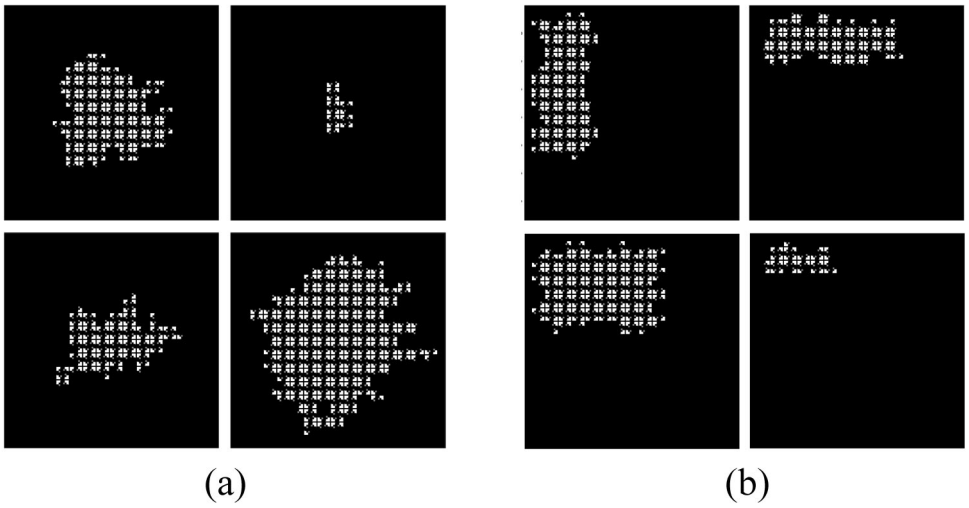

Figure 6.22 Example crystals generated that are (a) random and (b) rectangular shaped.

6.3.4.2 Diffraction Pattern Calculation

The crystals are zero-padded into a square array twice as large as their world array and their diffraction pattern calculated using the DFT. The intensities are averaged over the ensemble to calculate the averaged diffraction patterns. The B and $B3$ components for each crystal are separately calculated using the full-unit-cell part of the crystal and the incomplete-unit-cell part by themselves, respectively. The $B2$ component for each crystal is then calculated as $I^{B2}(\mathbf{u}) = I(\mathbf{u}) - I^{B}(\mathbf{u}) - I^{B3}(\mathbf{u})$. The averaged B, $B2$ and $B3$ components are obtained by averaging. The process is illustrated in Fig. 6.23.

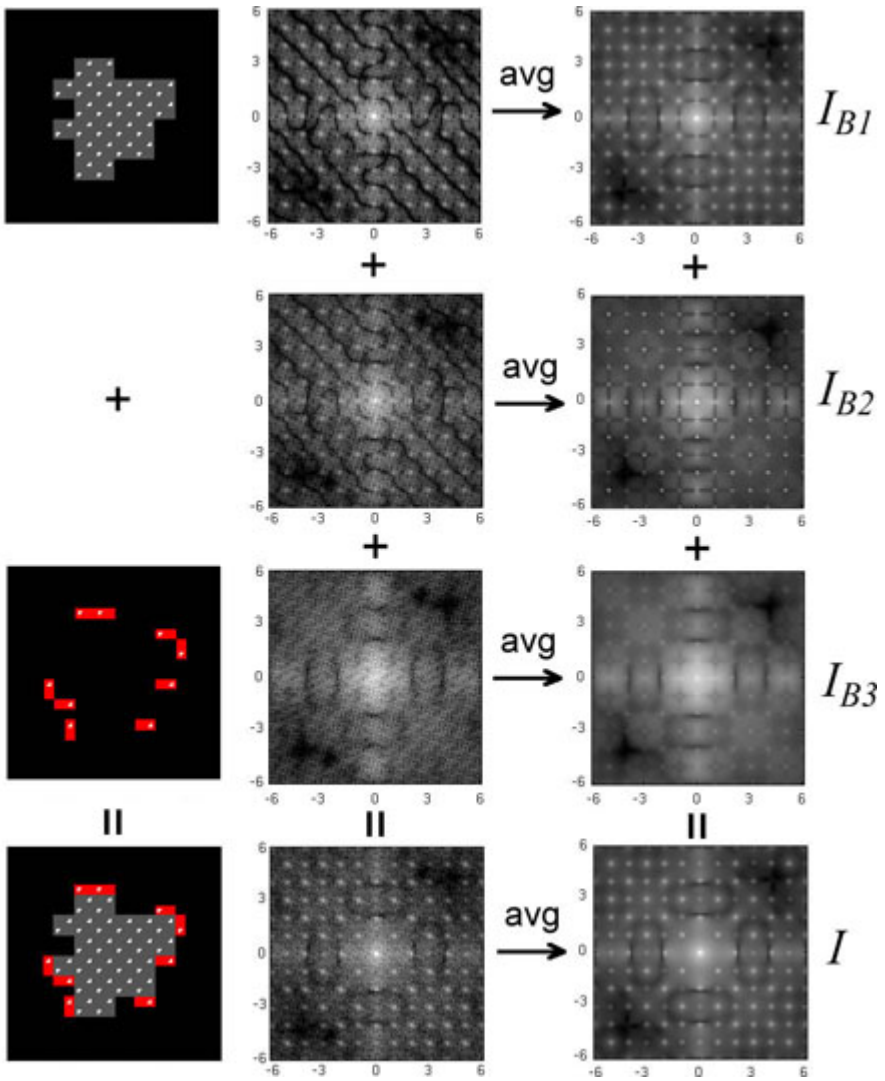

Figure 6.23 Pictorial representation of the process to calculate $I^{B2}(\mathbf{u})$, depicting the relation between each components. One instance crystal from the collection of crystals generated in the ensemble is shown in real space (left column) and its diffracted intensity in Fourier space (centre column). The averaged diffracted intensity from the entire ensemble is shown in the right column. The particular incomplete unit cell sites for this crystal are shown in red.

6.3.4.3 Simulation Results

Ensembles of 10,000 nanocrystals of the randomly-shaped and rectangular-shaped type were generated as described above. Rectangular crystals were generated with mean edge lengths equal to $\bar{N} = 5$ unit cells and standard deviation $\sigma_N = 1.67$ unit cells. Random shaped crystals were generated with mean size $\bar{N} = 25$ unit cells and standard deviation of $\sigma_N = 8$ unit cells. Ensemble-averaged diffraction patterns were calculated for both sets of crystals and the results are shown in Figs. 6.24 to 6.27. The three diffraction components and the total diffracted intensities for both crystal types are shown in Fig. 6.24. Each component is normalised to its peak value in the figure in order to clearly show its form, and so the $B2$ and $B3$ components are relatively weaker than they appear. The $B2$ component is negative and so its absolute value is displayed in Fig. 6.24. The Bragg component peaks at the reciprocal lattice points, and the $B2$ component shows streaks between the reciprocal lattice points, particularly evident for the rectangular crystals. The $B3$ component is more diffuse in nature but also shows some streaking for the case of rectangular crystals. A zoomed-in version of the $B2$ and $B3$ diffraction patterns is shown in Fig. 6.25, which shows that the $B2$ component vanishes and the $B3$ component is small, at the reciprocal lattice points, as anticipated in Sec. 6.3.3.

To study the nature of the diffraction in more detail, plots of the diffracted intensity along lines in reciprocal space are shown for the rectangular and random crystal ensembles in Figs. 6.26 and 6.27, respectively. Since the modulating effect of the molecular transform and its typically rapid fall-off at large distances wawy from the origin in reciprocal space tend to suppress the diffraction at high resolution, the intensity components are divided by $\langle |F_j(\mathbf{u})|^2 \rangle_j$ before being plotted in order to show them more clearly.

For the rectangular crystals, the diffracted intensity along the lines $u_2 = 0, 1/4, 1/2$ is shown in Fig. 6.26. Inspection of the figure shows that, as noted above, the $B2$ component is negative. This is likely related to the negative constant term in $Q^2_{B2}(\mathbf{u})$ noted in Sec. 6.3.3. The overall amplitudes of the $B2$ and $B3$ components are similar, with the $B2$ component being about twice that of the B3 component, which is consistent with Eq. (6.36). Inspection of Fig. 6.26(a) shows that the $B2$ component is zero at the reciprocal lattice points, and the $B3$ component is small. The B component dominates around the Bragg reflections but the other components are comparable to the Bragg component between the Bragg reflections. The diffracted intensity along the lines $u_2 = 1/4$ and $u_2 = 1/2$ is shown in Figs. 6.26(b) and (c), respectively, the latter being on a line midway between the reciprocal lattice points. As one moves away from the reciprocal lattice lines for rectangular crystals, the $B2$ component becomes larger (more negative) and cancels out some of the B component, and the $B3$ component dominates.

For the random-shaped crystals, the diffraction has similar characteristics but is

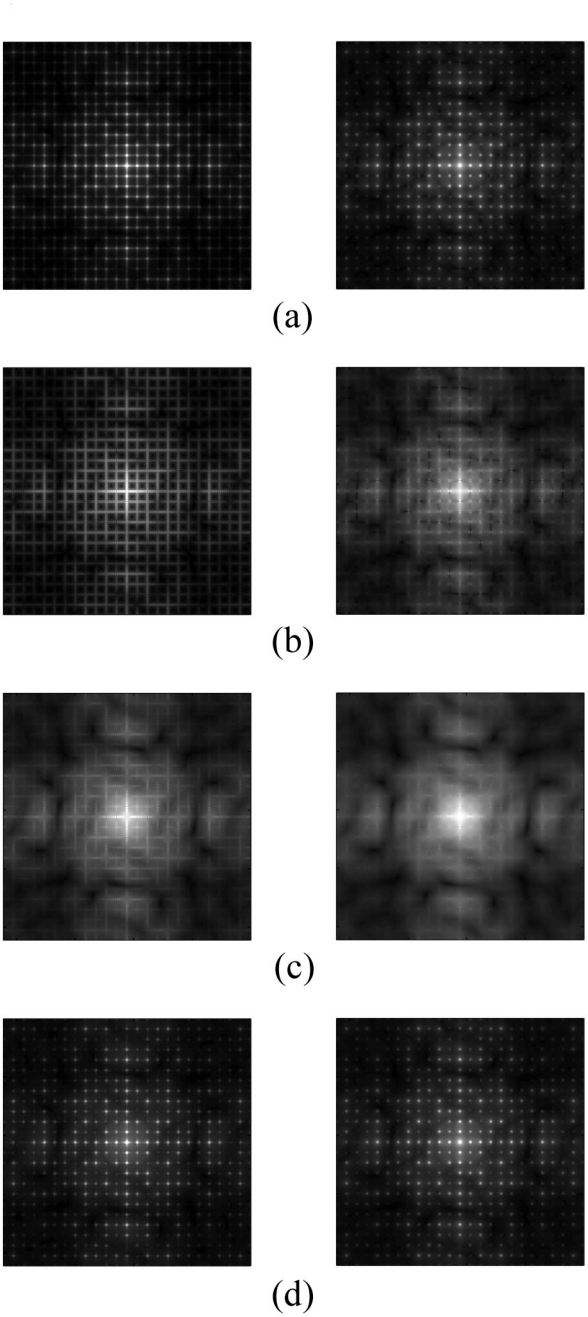

Figure 6.24 Ensemble averaged diffraction pattern components (a) B, (b) B2, (c) B3, and (d) the total diffraction, for rectangular (left) and random shaped (right) crystals with a mean of 25 unit cells as described in the text. Diffraction patterns are displayed as the fourth root of the intensity, and each pattern is normalised and thresholded, in order to clearly show the weak intensities. Increasing intensity is from black to white.

qualitatively different, as shown in Fig. 6.27. The behaviour on lines through the reciprocal lattice points is similar to the case of rectangular crystals. However, between the reciprocal lattice lines, the $B3$ component does not dominate over the Bragg component as much, and there is not as much cancellation between the B and $B2$ components.

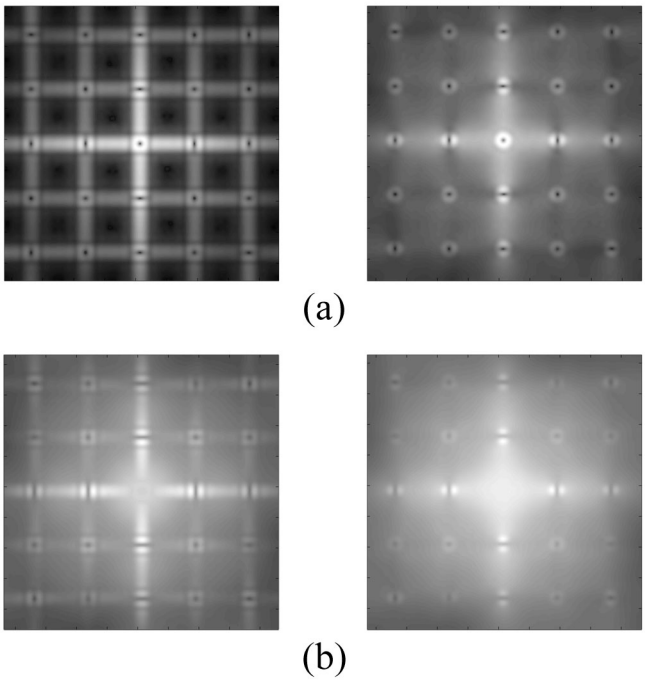

Figure 6.25 Zoomed-in views of the (a) $B2$ and (b) $B3$ components for rectangular (left) and random shaped crystals (right) in Figs. 6.24(b) and 6.24(c) around the origin, showing that the $B2$ component vanishes, and the $B3$ component is small, at the reciprocal lattice points.

Overall, the diffraction between the reciprocal lattice points is smoother and more diffuse with broader peaks for the random shaped crystals than for the rectangular crystals.

Note that in Figs. 6.26 and 6.27, the effect of normalising by $\langle|F_j(\mathbf{u})|^2\rangle_j$ is to show $Q^2_B(\mathbf{u})$ for the B component, but for the $B2$ and $B3$ components, the effect is to show $Q^2_{B2}(\mathbf{u})$ and $Q^2_{B3}(\mathbf{u})$ modulated by $2Re\big(\langle F_j(\mathbf{u})P^*_r(\mathbf{u})\rangle_{jr}\big)/\langle|F_j(\mathbf{u})|^2\rangle_j$ and $\langle P_r(\mathbf{u})P^*_{r'}(\mathbf{u})\rangle_{rr'}/\langle|F_j(\mathbf{u})|^2\rangle_j$, respectively. This highlights the effect of the different modulating factors in Eq. (6.30) for the $B2$ and $B3$ components.

Diffraction patterns were also calculated for larger crystals with mean edge lengths of 14 unit cells and standard deviation of 4 unit cells for the rectangular crystals, and for random shaped crystals, with a mean of 200 unit cells and a standard deviation of 70 unit cells. The corresponding diffraction profiles are shown in Figs. 6.28 and 6.29, respectively. For the rectangular crystals, the diffraction peaks are narrower than for the smaller crystals. The $B3$ component is weaker near the reciprocal lattice lines, but still dominates between those lines and has the same overall shape as for the smaller crystals. For the random-shaped crystals, the peaks are narrower along the reciprocal lattice lines and the Bragg-like components are weaker than for the smaller crystals. However, off the reciprocal lattice lines all components have strikingly similar peak widths and relative magnitudes to the case for smaller crystals.

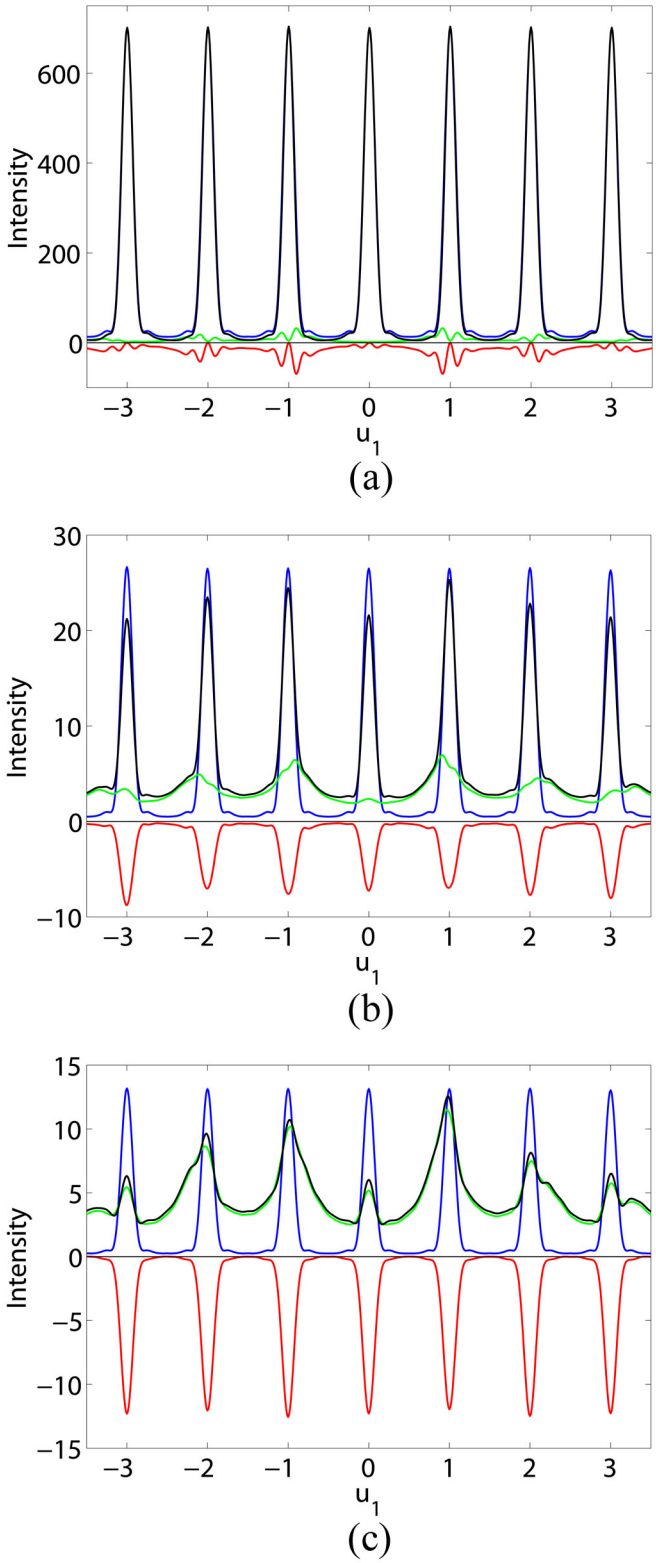

Figure 6.26 Diffracted intensity components divided by $\langle |F_j(\mathbf{u})|^2 \rangle_j$ for rectangular crystals with a mean of 25 unit cells: B (blue), $B2$ (red), $B3$ (green) and total (black) as a function of u_1 for (a) $u_2 = 0$, (b) $u_2 = 1/4$ and (c) $u_2 = 1/2$.

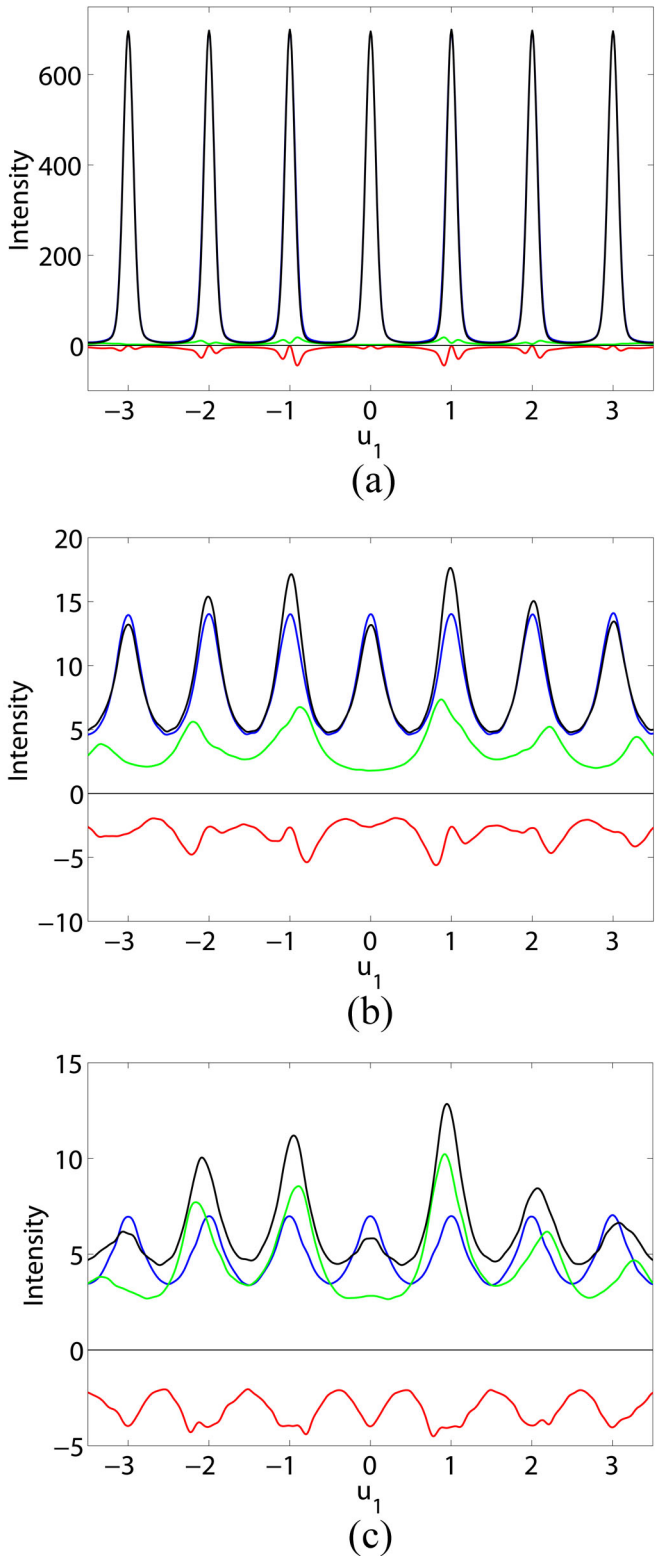

Figure 6.27 Diffracted intensity components divided by $\langle |F_j(\mathbf{u})|^2 \rangle_j$ for random shaped crystals with a mean of 25 unit cells: B (blue), $B2$ (red), $B3$ (green) and total (black) as a function of u_1 for (a) $u_2 = 0$, (b) $u_2 = 1/4$ and (c) $u_2 = 1/2$.

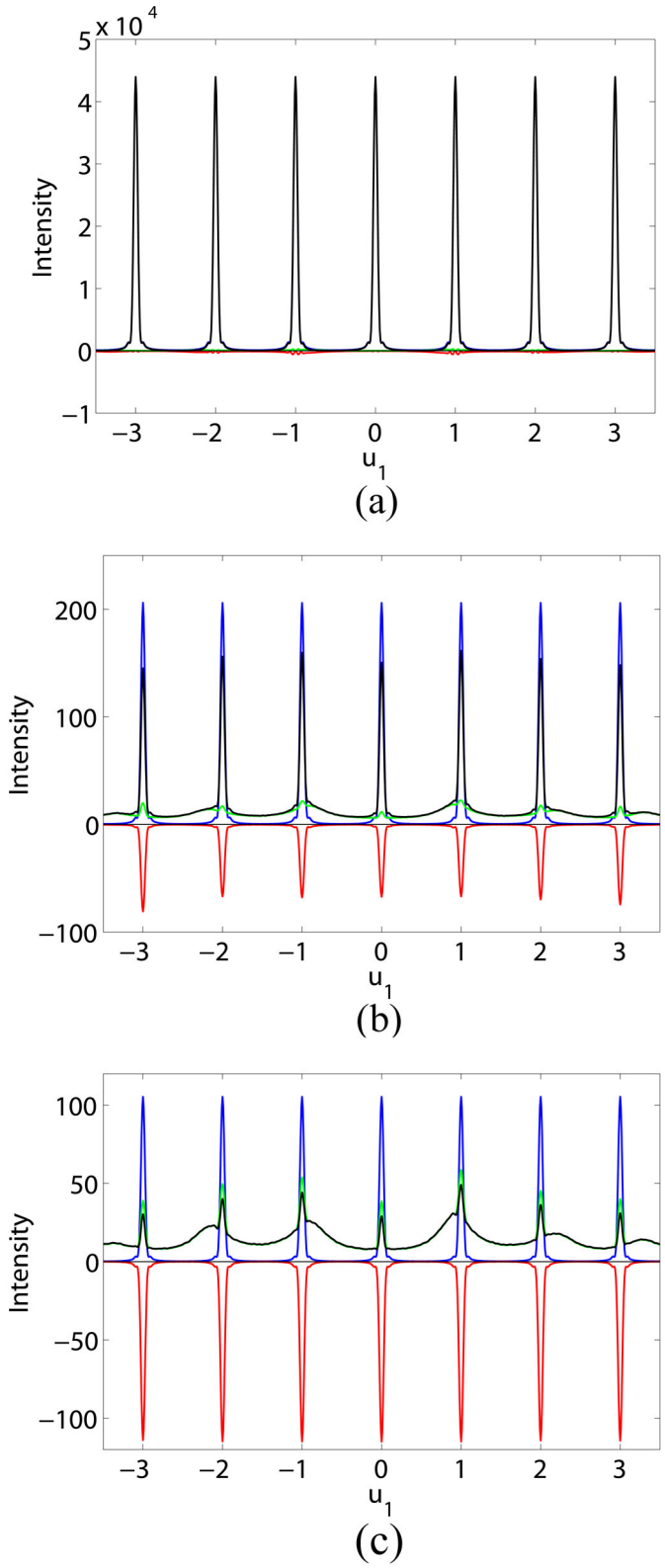

Figure 6.28 Diffracted intensity components divided by $\langle |F_j(\mathbf{u})|^2 \rangle_j$ for rectangular crystals with a mean of 200 unit cells: B (blue), $B2$ (red), $B3$ (green) and total (black) as a function of u_1 for (a) $u_2 = 0$, (b) $u_2 = 1/4$ and (c) $u_2 = 1/2$.

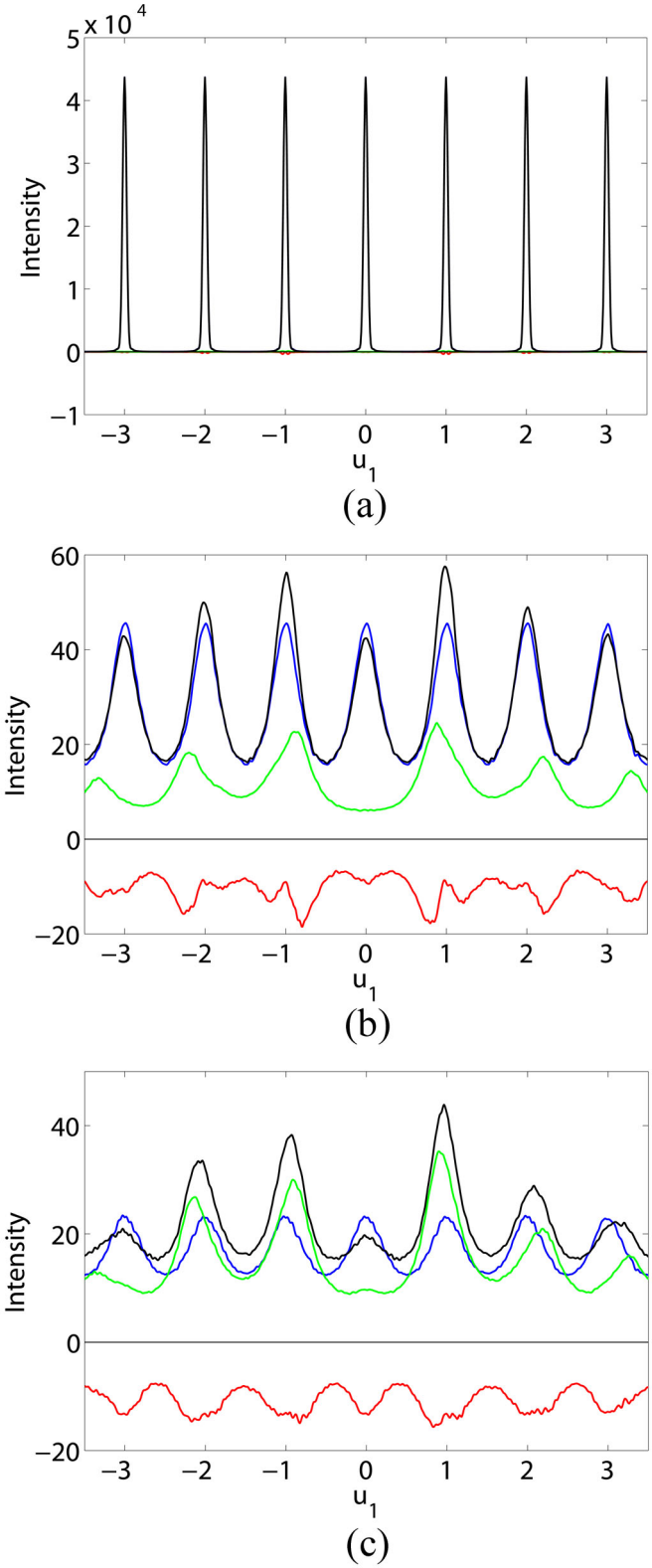

Figure 6.29 Diffracted intensity components divided by $\langle |F_j(\mathbf{u})|^2 \rangle_j$ for random shaped crystals with a mean of 200 unit cells: B (blue), $B2$ (red), $B3$ (green) and total (black) as a function of u_1 for (a) $u_2 = 0$, (b) $u_2 = 1/4$ and (c) $u_2 = 1/2$.

6.4 MINIMISING INCOMPLETE UNIT CELLS

It is worth noting that the constructive procedure described above will not minimise the incomplete-unit-cell part of the crystal in every case. There will be some cases where, after constructing the crystal, a reassignment of the *kind* of complete unit cell will allow the full-unit-cell part of the crystal to be larger as was illustrated in Figs. 6.12(a) and (b). This can be realised by recalling the form of the electron density of the crystal given by Eq. (6.26), stated here again for ease of reference,

$$g(\mathbf{x}) = \sum_{\mathbf{m}} f_j(\mathbf{x} - \mathrm{L}\mathbf{m}) + p_j(\mathbf{x}). \tag{6.37}$$

Because different unit cell definitions result in different molecular terminations on the crystal surface (both f and p are indexed by j), the unit cell type that minimises the surface component $p_j(\mathbf{x})$ will be the most desirable. Since this can always be achieved by defining the unit cell to be the crystal itself, a further constraint that should be imposed on this minimisation problem is that the period of the crystal must be greater than one, i.e., there must be more than one copy of the unit cell in the crystal. The minimisation can also be restricted to a set of complete unit cells $\{f_j(\mathbf{x})\}$ that has already been decided. An appropriate set of complete unit cells to consider is those that are the same size and shape as the basic unit cell of the lattice at hand and are wrapped translations of each other that contain full molecules, as shown in Fig. 6.14.

In practice, the probabilities of the occurrence of different surface structures would be complicated and depend on the structure of the molecule and the nature of the bonds between its atoms. A useful definition of the different kinds of complete unit cell may thus need to be considered on a case-by-case basis, and would depend on the space group and molecular arrangement at hand.

6.5 CONCLUDING REMARKS

The manner in which unit cells are terminated on the crystal surface becomes important when the crystals are small and when diffraction between the Bragg peaks is of interest. If there is more than one molecule in the unit cell, as is usually the case, then different crystals can be described by different possible unit cells that are translationally wrapped versions of each other and contain whole molecules. This is essentially the non-uniqueness of the period in a periodic function.

Conventional crystallography does not suffer from the effect of incomplete unit cells because (1) the crystals used are large and so the effects of their surface on the diffraction are small, and (2) only diffraction information at the Bragg peaks is used, which is independent of the details on the crystal edge since Bragg reflections represent the coefficients of a Fourier series which assumes the crystal is infinitely periodic, i.e.

with no edge.

A nanocrystal can be represented as the sum of two parts: a part that consists of identical, complete unit cells, and a part that consists of incomplete, or partial, unit cells occurring on the surface of the crystal. Each incomplete unit cell has fewer than the full number of molecules in the unit cell. The ensemble-averaged diffraction from nanocrystals described in this way consists of three components: the usual Bragg component (denoted the B component) due to the full-unit-cell part of the crystals, and two Bragg-like components; one, denoted the $B2$ component, that is due to interference between the full-unit-cell and incomplete-unit-cell parts of the crystals, and the other, the $B3$ component, due to interference within the incomplete-unit-cell part of the crystals themselves. The reason why all three components of the averaged diffracted intensity are attached the name "Bragg" is that all three show evidence of Bragg-like behaviour, peaking at the Bragg positions, i.e. none of them are completely diffuse. These are seen in the results of the simulations conducted in 2-D. There is some flexibility in the definition of the Bragg and Bragg-like components in terms of the unit cell used and, since the Bragg-like components can be treated as a perturbation on the Bragg component, the problem can be formulated in such a way as to minimise the overall magnitude of the Bragg-like component. By defining the full-unit-cell part of the crystals such that the incomplete-unit-cell part is minimised, a more useful description of the ensemble averaged diffraction can be obtained.

In 1-D, the averaged diffraction from crystals modelled in this fashion was found to have a closed-form solution. In higher dimensions, simulation was used to gauge the overall picture of the averaged diffracted intensity due to the multitude of crystal shapes possible. The simulation involved growing crystals probabilistically via a random-walk-type process where each unit cell has equal probability of being attached onto any one side of the crystal. Thus if an empty cell site is adjacent to two occupied sites, the probability of that empty site being occupied becomes twice as large. This tendency for cells to bunch leads to clustering as the number of unit cells become large and so does not model actual crystal growth in the real world very well. Real crystals will generally tend to have facets and planar edges rather than forming a spherical clump. However for small crystals, facets may not be so prominent even in practice, and so the highly simplistic model employed may give adequate crystal shapes for crystals used in nanocrystallography. In any case, the procedure provides an interesting and seemingly fundamental way of generating ensembles of shapes to allow study of the averaged diffraction from small crystals.

The overall intensities of the $B2$ and $B3$ components are both of order $\bar{N}^{-1/3}$ compared to the Bragg component for 3-D crystals. However, although the Bragg component dominates near the Bragg reflections, for highly faceted crystals the $B3$ component dominates between the reciprocal lattice lines (or planes in 3-D). For weakly faceted crystals, the three components are of more similar magnitude away from the

Bragg reflections. The results show that the presence of incomplete unit cells on the crystal surface affects the intensity distribution *between* the Bragg reflections and this may influence proposals for direct phasing. The results also show that the incomplete unit cells have a small effect on the intensities *at* the Bragg reflections through the $B3$ component.

Chapter 7

PHASE RETRIEVAL OF NANOCRYSTAL DIFFRACTION WITH MULTIPLE UNIT CELLS

7.1 INTRODUCTION

It was shown in Chapter 6 that averaging diffraction patterns from many crystals with different surface terminations in an SFX experiment generates a diffraction data set that, to a first approximation, is averaged over a set of complete unit cells. The method of phase retrieval described in Chapter 5 where the continuous Fourier intensity of one unit cell is obtained by dividing the averaged diffraction by the averaged squared shape transform $Q^2(\mathbf{u})$ therefore gives access in actuality to only the averaged Fourier intensity from some set of distinct unit cells. The data available for phase retrieval are then (1) the averaged Fourier intensity, (2) the support of the unit cells and (3) the symmetry relationships within and between the different kinds of unit cell. The problem posed in this way gives rise to a generalised version of the phase problem in which it is desired to reconstruct a small set of objects from measurements of only the average (or sum) of their Fourier intensities.

This problem is addressed in two parts. The first considers the problem of reconstructing a small set of independent objects from only the average of all the Fourier intensity of the objects in that set. The second part considers the same problem but for the case of dependent objects where the objects are related to each other - which is the case that arises with multiple unit cell definitions in SFX.

7.2 PHASE RETRIEVAL WITH MULTIPLE INDEPENDENT OBJECTS

Let there be a set of J objects $\{f_1(\mathbf{x}), \cdots, f_J(\mathbf{x})\}$ where the jth object of the set is denoted $f_j(\mathbf{x})$ and has support $\mathcal{S}_j$, autocorrelation $A_j(\mathbf{x})$ and Fourier transform $F_j(\mathbf{u})$. The autocorrelation has support $\mathcal{A}_j$ and the averaged, or equivalently, summed Fourier

intensity $I(\mathbf{u})$ is given by

$$I(\mathbf{u}) = \sum_{j=1}^{J} |F_j(\mathbf{u})|^2. \tag{7.1}$$

The problem is to reconstruct the set of objects $\{f_j(\mathbf{x})\}$ given $I(\mathbf{u})$.

7.2.1 Uniqueness - Constraint Ratio

Averaging the Fourier intensities together reduces the amount of information compared to that when individual intensities are accessible. Given this loss of information, the problem can have a unique solution only if there is some redundancy in the phase problem for a single object. This is indeed the case and the amount of information can be quantified by the constraint ratio introduced in Sec. 3.3.3.5.

7.2.1.1 Single Object

Recall from Sec. 3.3.3.5, uniqueness of the phase problem can be quantified using the constraint ratio, Ω, defined as the ratio of the number of independent data and the number of independent free parameters in the object [Elser and Millane 2008]. Uniqueness of the reconstruction problem requires the number of independent data to exceed the number of free parameters in the object. In phase retrieval, the free parameters correspond to the unknown values of the samples of the object density and the data are given by the known values of the independent samples of the autocorrelation of that density. The number of samples in an object is equivalent to the number of samples in its support, and so the constraint ratio for the jth object in the set is given by

$$\Omega_j = \frac{|\mathcal{A}_j|}{2|\mathcal{S}_j|}\,, \tag{7.2}$$

where $|\cdot|$ is the number of samples. Uniqueness of the phase problem then corresponds to $\Omega > 1$. The case $\Omega = 1$ is marginal in the sense that it may admit a one-parameter family of solutions, and some additional information is required to restore uniqueness. For one object in two or more dimensions, the constraint ratio is bounded below by 2^{D-1} where D is the dimension of the object as mentioned in Sec. 3.3.3.5. The phase problem is thus overdetermined and some loss of information can potentially be tolerated in the problem at hand with reconstructing multiple objects.

7.2.1.2 Discrete Objects

Equation (7.2) for the calculation of Ω above applies to continuous, as opposed to sampled, objects, as is necessary when the objects are discretised in a computer. For sampled objects, the value of the constraint ratio is smaller than for continuous objects with supports of the same shape. This can be seen by noting for example that a 2-D

rectangular object of $N \times N$ samples has an autocorrelation of $(2N-1) \times (2N-1)$ samples, as opposed to $2N \times 2N$ samples in the continuous case, due to the way the convolution of a discrete signal is calculated. The loss of one sample on each side therefore leads to a slightly smaller value of Ω for discrete objects.

To quantify this, denote the constraint ratio calculated for discrete objects by Ω_{D}, the ratio of Ω_{D} to Ω for a D-dimensional hyper-cubic object with a total of Q samples is,

$$\begin{aligned}\frac{\Omega_{\mathrm{D}}}{\Omega} &= \frac{\left(2Q^{\frac{1}{D}}-1\right)^{D}/(2Q)}{\left(2Q^{\frac{1}{D}}\right)^{D}/(2Q)} \\ &= \left(1-\frac{1}{2}Q^{-\frac{1}{D}}\right)^{D} \\ &= 1-\frac{1}{2}DQ^{-\frac{1}{D}}+O(Q^{-\frac{2}{D}}),\end{aligned} \tag{7.3}$$

where the binomial expansion is utilised for the last equality. Therefore, for a D-dimensional discrete object with a total of Q samples, the value of Ω calculated on the basis of a continuous object is in error by $O(Q^{-1/D})$. This effect is small except for very small objects (objects with a small number of samples), with $\Omega_{\mathrm{D}} \to \Omega$ as $Q \to \infty$. A plot of Ω_{D} versus the first order term $1-DQ^{-\frac{1}{D}}$ for the 3-D cubic case is shown in Fig. 7.1. Extrapolating to $1-DQ^{-\frac{1}{D}}=1$ shows $\Omega_{\mathrm{D}} \to 4$ when $Q \to \infty$ as expected. The continuous approximation to Ω will be used henceforth unless explicitly mentioned otherwise.

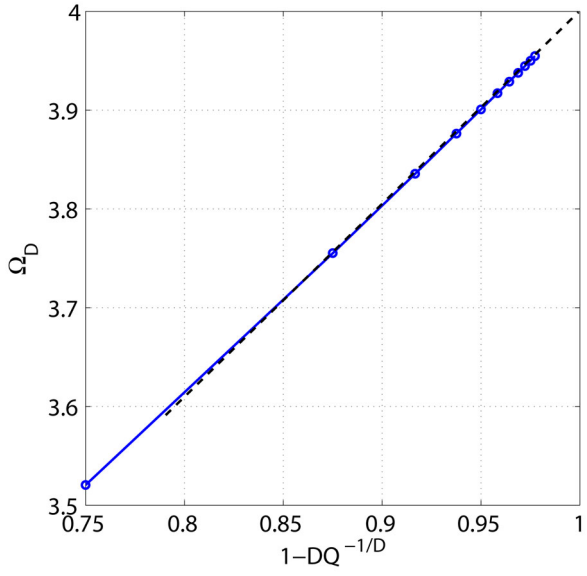

Figure 7.1 The constraint ratio calculated for discretised cubes of increasing side lengths, or equivalently, of increasing number of samples per side with a fixed side length, measured by $1-DQ^{-\frac{1}{D}}$. The dotted line shows the extrapolation of Ω_{D} to get Ω which occurs when $Q \to \infty$ or $1-DQ^{-\frac{1}{D}} \to 1$.

7.2.1.3 Multiple Objects

Consider now the constraint ratio for multiple objects. The number of free parameters for a set of J independent objects is equal to the total number of samples from all the objects, given by

$$|\mathcal{S}| = \sum_{j=1}^{J} |\mathcal{S}_j|. \tag{7.4}$$

The inverse Fourier transform of $I(\mathbf{u})$, the sum of the Fourier intensity from all the objects, is the sum of the autocorrelations of all the objects, denoted by $A(\mathbf{x})$, i.e. $A(\mathbf{x}) = \sum_{j=1}^{J} A_j(\mathbf{x})$. The support of $A(\mathbf{x})$, denoted by $\mathcal{A}$, is therefore the union of all the supports of the autocorrelation of the individual objects, $\mathcal{A}_j$, giving

$$|\mathcal{A}| = \left| \bigcup_{j=1}^{J} \mathcal{A}_j \right|. \tag{7.5}$$

The number of independent samples of information contained in $I(\mathbf{u})$ is therefore equal to $|\mathcal{A}|/2$ and the constraint ratio Ω for the averaged intensity phase problem is then given by

$$\Omega = \frac{|\mathcal{A}|}{2|\mathcal{S}|} \, . \tag{7.6}$$

Knowing the supports $\mathcal{S}_j$, uniqueness for a particular problem can be determined by calculating Ω and checking that $\Omega > 1$. This will depend on the shapes of the individual object supports $\mathcal{S}_j$ and their dimensionality.

The constraint ratio for multiple objects is illustrated by a simple example as follows. Consider two rectangular objects with supports $a_1 \times b_1$ and $a_2 \times b_2$ so that $|\mathcal{S}| = a_1 b_1 + a_2 b_2$ where the a_i and b_i are real constants. The autocorrelation of these two objects are also rectangular and has a support of size $2a_1 \times 2b_1$ and $2a_2 \times 2b_2$. There are two distinct cases to consider, depending on the relative dimensions of the two rectangles. Case A, one autocorrelation support fits completely inside the other, and case B, one autocorrelation support does not fit completely inside the other. Both cases are illustrated in Fig. 7.2.

Case A occurs if either $a_1 \leq a_2$ and $b_1 \leq b_2$ or $a_1 \geq a_2$ and $b_1 \geq b_2$. Case B occurs if either $a_1 < a_2$ and $b_1 > b_2$ or $a_1 > a_2$ and $b_1 < b_2$. Calculation of the constraint ratio then gives

$$\Omega = \begin{cases} \frac{2}{1+r_a r_b} & \text{Case A} \\ 2\left(1 - \frac{r_a r_b}{r_a + r_b}\right) & \text{Case B,} \end{cases} \tag{7.7}$$

where

$$r_a = \frac{\min(a_1, a_2)}{\max(a_1, a_2)} \tag{7.8}$$

and

$$r_b = \frac{\min(b_1, b_2)}{\max(b_1, b_2)}. \tag{7.9}$$

Note that r_a and r_b are between 0 and 1. For both cases A and B, $1 \leq \Omega < 2$ with $\Omega = 1$ if the two supports are the same (r_a, r_b tends to 1). For case A, Ω tends to 2 if one of the objects become very small. For case B, Ω tends to 2 if one or both of the widths of the two objects are very different, i.e. $r_a \to 0$ or $r_b \to 0$. Therefore, for a set of rectangular supports, the solution to the phase problem is unique as long as the object supports are different.

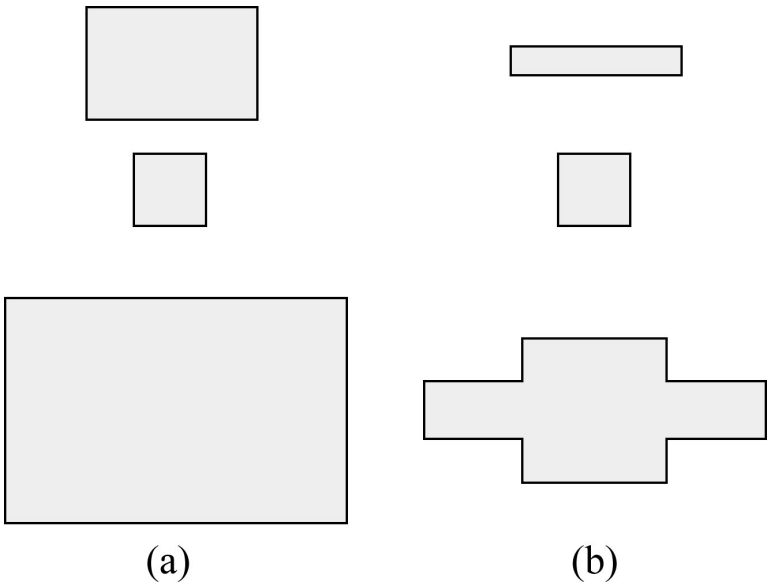

Figure 7.2 Object supports (top two rows) and respective autocorrelation supports (bottom row) for rectangular objects for (a) Case A and (b) Case B.

A special case occurs when all J objects have an identical support. Since $f_j(\mathbf{x})$ have the same support for all j, the supports of the corresponding autocorrelations $A_j(\mathbf{x})$ and their sum, $A(\mathbf{x})$, are also all identical. Therefore, by the union operation in Eq. (7.5), $|\mathcal{A}| = |\mathcal{A}_j|$ for all j, and the number of independent data remains the same as for a single object. The number of object parameters have however increased by a factor of J. The constraint ratio for the averaged data from J objects with the same support is then given by $\Omega = \Omega_j/J$ where Ω_j is the constraint ratio for a single object.

A unique solution to the phase problem for J objects with identical supports is therefore expected if $\Omega > J$. The largest value of J that can be tolerated depends on the dimensionality of the problem and the nature of the object supports. For example, for a 2-D rectangular support it is required that $J < 2$, for a 3-D cuboid support $J < 4$, and for a 2-D triangular support $J < 3$ as given by the constraint ratio for these single object supports quoted in Sec. 3.3.3.5. Any additional constraint is likely to reduce these strict inequalities ($<$) to non-strict inequalities ($\leq$).

In summary, the phase problem for averaged Fourier intensity data from multiple

objects has a unique solution if $\Omega > 1$. If some of the objects have the same support and some have different supports, uniqueness of the phase problem can still be easily determined by calculating the constraint ratio and ensuring that it is greater than one.

7.2.2 Uniqueness - Fundamental Ambiguity

If the supports of all J objects are the same, then additional ambiguities exist. As noted in Sec. 3.3.3.5, the phase problem for a single object has fundamental ambiguities in that $f(\mathbf{x}-\mathbf{a})\exp(ib)$ and $f^*(-\mathbf{x}-\mathbf{a})\exp(ib)$, have the same Fourier intensity as $f(\mathbf{x})$, where $\mathbf{a}$ is a constant vector and b is a real constant. These ambiguities are also present for the averaged intensity problem for each of the objects $f_j(\mathbf{x})$. Furthermore, ambiguities related to the mixing of different objects also exist for the averaged Fourier intensity phase problem with identical supports, which can be shown as follows.

Writing the set of objects $\{f_j\}$ and the set of their Fourier transforms $\{F_j\}$ as vectors $\mathbf{f}(\mathbf{x}) = [f_1, \cdots, f_J]^T$ and $\mathbf{F}(\mathbf{u}) = [F_1, \cdots, F_J]^T$, the Fourier data that is available can be expressed as

$$I_F(\mathbf{u}) = \sum_{j=1}^{J} |F_j|^2 = \mathbf{F}^T\mathbf{F}^*, \tag{7.10}$$

where the complex conjugate applies individually to each F_j in $\mathbf{F}$. An ambiguity can arise if there are other sets of images with the same support that satisfy Eq. (7.10).

Consider a set of J objects $\{g_j\}$ where each object of the set is a linear combination of $f_j(\mathbf{x})$ and its inverted version $f_j(-\mathbf{x})$, i.e.,

$$\begin{aligned} g_1(\mathbf{x}) &= a_{11}f_1(\mathbf{x}) + \cdots + a_{1J}f_J(\mathbf{x}) + b_{11}f_1(-\mathbf{x}) + \cdots + b_{1J}f_J(-\mathbf{x}) \\ &\vdots \\ g_J(\mathbf{x}) &= a_{J1}f_1(\mathbf{x}) + \cdots + a_{JJ}f_J(\mathbf{x}) + b_{J1}f_1(-\mathbf{x}) + \cdots + b_{JJ}f_J(-\mathbf{x}), \end{aligned} \tag{7.11}$$

where a_{ij} and b_{ij} are real constants. Writing the set of all objects $\{g_j\}$ as the vector $\mathbf{g}(\mathbf{x}) = [g_1, \cdots, g_J]^T$, Eq. (7.11) can be expressed as

$$\mathbf{g}(\mathbf{x}) = \mathrm{A}\mathbf{f}(\mathbf{x}) + \mathrm{B}\mathbf{f}(-\mathbf{x}), \tag{7.12}$$

where the entries of the $J \times J$ matrices A and B are the coefficients a_{jj} and b_{jj} of the linear combinations that generate the set of objects $\mathbf{g}(\mathbf{x})$, respectively. Assuming the set of objects $\mathbf{f}(\mathbf{x})$ is real, the Fourier transform of $\mathbf{g}(\mathbf{x})$ is

$$\begin{aligned} \mathbf{G}(\mathbf{u}) &= \mathrm{A}\mathbf{F}(\mathbf{u}) + \mathrm{B}\mathbf{F}(-\mathbf{u}) \\ &= \mathrm{A}\mathbf{F}(\mathbf{u}) + \mathrm{B}\mathbf{F}^*(\mathbf{u}). \end{aligned} \tag{7.13}$$

Dropping the variable $\mathbf{u}$ for clarity,

$$\mathbf{G} = \mathrm{A}\mathbf{F} + \mathrm{B}\mathbf{F}^*. \tag{7.14}$$

The summed Fourier intensity of the set of all objects in $\mathbf{g}$ is then

$$I_G(\mathbf{u}) = \sum_{j=1}^{J} |G_j|^2 = \mathbf{G}^T\mathbf{G}^*. \tag{7.15}$$

Substitution of Eq. (7.14) into Eq. (7.15) leads to

$$\begin{aligned} I_G(\mathbf{u}) &= \mathbf{G}^T\mathbf{G}^* \\ &= (\mathrm{A}\mathbf{F} + \mathrm{B}\mathbf{F}^*)^T (\mathrm{A}\mathbf{F} + \mathrm{B}\mathbf{F}^*)^* \\ &= \left(\mathbf{F}^T\mathrm{A}^T + \mathbf{F}^{*T}\mathrm{B}^T\right)\left(\mathrm{A}\mathbf{F}^* + \mathrm{B}\mathbf{F}\right) \\ &= \mathbf{F}^T\mathrm{A}^T\mathrm{A}\mathbf{F}^* + \mathbf{F}^{*T}\mathrm{B}^T\mathrm{B}\mathbf{F} + \mathbf{F}^T\mathrm{A}^T\mathrm{B}\mathbf{F} + \mathbf{F}^{*T}\mathrm{B}^T\mathrm{A}\mathbf{F}^* \\ &= \mathbf{F}^T\mathrm{A}^T\mathrm{A}\mathbf{F}^* + \mathbf{F}^T\mathrm{B}^T\mathrm{B}\mathbf{F}^* + \mathbf{F}^T\mathrm{A}^T\mathrm{B}\mathbf{F} + \mathbf{F}^T\mathrm{B}^T\mathrm{A}\mathbf{F} \\ &= \mathbf{F}^T(\mathrm{A}^T\mathrm{A} + \mathrm{B}^T\mathrm{B})\mathbf{F}^* + \mathbf{F}^T(\mathrm{A}^T\mathrm{B} + \mathrm{B}^T\mathrm{A})\mathbf{F}, \end{aligned} \tag{7.16}$$

keeping in mind that the coefficients of A and B are real so that $\mathbf{F}^{*T}\mathrm{B}^T\mathrm{B}\mathbf{F} = \mathbf{F}^T\mathrm{B}^T\mathrm{B}\mathbf{F}^*$ and $\mathbf{F}^{*T}\mathrm{B}^T\mathrm{A}\mathbf{F}^* = \mathbf{F}^T\mathrm{B}^T\mathrm{A}\mathbf{F}$.

For $I_G(\mathbf{u})$ to equal $I_F(\mathbf{u})$ and therefore an ambiguity to occur, the coefficient matrices A and B must thus satisfy

$$\mathrm{A}^T\mathrm{A} + \mathrm{B}^T\mathrm{B} = \mathrm{I} \tag{7.17}$$

$$\mathrm{A}^T\mathrm{B} + \mathrm{B}^T\mathrm{A} = 0, \tag{7.18}$$

where I and 0 are the $J \times J$ identity and zero matrices respectively. Conditions (7.17) and (7.18) together say that A + B is orthogonal and $\mathrm{A}^T\mathrm{B}$ is skew symmetric. For a set of identical objects $\mathbf{f}$ that have a support shape that is non-centrosymmetric, B = 0 because $\mathbf{f}(\mathbf{x})$ and $\mathbf{f}(-\mathbf{x})$ cannot additively combine to give a support shape that is the same as the one for either of themselves since they have different supports. The condition for non-uniqueness for this case then becomes $\mathrm{A}^T\mathrm{A} = \mathrm{I}$, i.e., A must be orthogonal.

The two conditions in Eqs. (7.17) and (7.18) can be written together as

$$\begin{bmatrix} \mathrm{I} & 0 \\ 0 & \mathrm{I} \end{bmatrix} = \begin{bmatrix} \mathrm{A}^T & \mathrm{B}^T \\ \mathrm{B}^T & \mathrm{A}^T \end{bmatrix} \begin{bmatrix} \mathrm{A} & \mathrm{B} \\ \mathrm{B} & \mathrm{A} \end{bmatrix} = \begin{bmatrix} \mathrm{A} & \mathrm{B} \\ \mathrm{B} & \mathrm{A} \end{bmatrix}^T \begin{bmatrix} \mathrm{A} & \mathrm{B} \\ \mathrm{B} & \mathrm{A} \end{bmatrix} = \mathrm{I}, \tag{7.19}$$

where the second equality results from the identity

$$\begin{bmatrix} \mathrm{A}_1 & \mathrm{A}_2 \\ \mathrm{A}_3 & \mathrm{A}_4 \end{bmatrix}^T = \begin{bmatrix} \mathrm{A}_1^T & \mathrm{A}_3^T \\ \mathrm{A}_2^T & \mathrm{A}_4^T \end{bmatrix} \tag{7.20}$$

for $\mathrm{A}_1, \cdots, \mathrm{A}_4$ being square matrices.

With $\mathbf{g}(\mathbf{x})$ being a linear combination of $\mathbf{f}(\mathbf{x})$ and $\mathbf{f}(-\mathbf{x})$, it means that $\mathbf{f}(\mathbf{x})$ can also be expressed as a linear combination of $\mathbf{g}(\mathbf{x})$ and $\mathbf{g}(-\mathbf{x})$. Using subscripts f and g to distinguish between the A and B coefficient matrices for $\mathbf{f}(\mathbf{x})$ and $\mathbf{g}(\mathbf{x})$, respectively, write

$$\mathbf{g}(\mathbf{x}) = \mathrm{A}_f \mathbf{f}(\mathbf{x}) + \mathrm{B}_f \mathbf{f}(-\mathbf{x}) \tag{7.21}$$

and

$$\mathbf{f}(\mathbf{x}) = \mathrm{A}_g \mathbf{g}(\mathbf{x}) + \mathrm{B}_g \mathbf{g}(-\mathbf{x}). \tag{7.22}$$

The coefficient matrices for $\mathbf{f}(\mathbf{x})$ and $\mathbf{g}(\mathbf{x})$ are then related by

$$\begin{aligned} \mathrm{A}_g &= \mathrm{A}_f(\mathrm{A}_f + \mathrm{B}_f)^{-1}(\mathrm{A}_f - \mathrm{B}_f)^{-1} \\ \mathrm{B}_g &= \mathrm{B}_f(\mathrm{B}_f + \mathrm{A}_f)^{-1}(\mathrm{B}_f - \mathrm{A}_f)^{-1} \end{aligned} \tag{7.23}$$

such that one set of coefficient matrices can be found from the other.

In summary, there are an infinite number of solutions that will satisfy the phase problem for a set of objects with identical supports. More specifically, in the set of $J \times J$ matrices, there is a $J(J-1)/2$-parameter family of orthogonal matrices, so there is correspondingly a $J(J-1)/2$-parameter family of solutions to the averaged intensity phase problem for J objects with identical supports. Therefore, even in the case $\Omega > 1$, the problem is highly non-unique in general. Note that for a set of objects with different supports, alternative objects generated as linear combinations of the objects and their centrosymmetric inversions from the set would have supports that are larger than those of the original objects. The supports of these objects are therefore inconsistent with the object support constraints so that the fundamental ambiguity as described above are excluded for a set of objects with non-identical supports.

7.2.2.1 Positive Objects

In many practical situations the object functions $f_j(\mathbf{x})$ are real and positive and these properties reduce the non-uniqueness of the phase retrieval problem. If the objects are positive and real, then the entries of the matrices A and B are either equal to $+1$, -1 or 0, in which case the nonuniqueness corresponds only to the fundamental ambiguities in phase retrieval for a single object, or some of the coefficients must take on general negative values ($\neq -1$). If the true objects $f_j(\mathbf{x})$ are positive real, then it is likely that these general negative coefficients will produce at least one object $g_j(\mathbf{x})$ that is

negative for some of its samples, and these solutions are thus excluded by the positivity constraint. It is therefore likely that the positivity constraint will eliminate many of the incorrect solutions. Furthermore, positivity is expected to exhibit better constraining on the solution if the objects have a wide range of sample values, since it is then more likely that a linear combination of the objects will produce a negative sample even for small negative coefficients.

7.2.3 Phase Retrieval

Iterative projection algorithms are adapted here for phase retrieval for multiple objects. The objects $f_j(\mathbf{x})$ for all $j = 1, \cdots, J$ is represented as a vector in vector space. Assume that each object is sampled such that they each have a total of Q samples. The vector space is then of dimension JQ. The phase retrieval problem is formulated as finding a point in this vector space that is in the intersection of two constraint sets. One constraint set contains all objects that satisfy the real space constraints, which in this case is the required support of each object and possibly positivity. The other constraint set contains the set of all J objects whose summed Fourier intensities are equal to the measured intensity $I(\mathbf{u})$.

The support projection P_S is implemented in the usual way by setting the value of the samples that are outside the support of each object to zero and leaving the values of the samples inside the support unchanged. Similarly, the positivity projection, if applied, is implemented by setting negative sample values in each object to zero. Consider now the Fourier space projection P_M. This projection corresponds to making the smallest change to the current set of J objects such that their corresponding Fourier intensities sum to the measured function $I(\mathbf{u})$ given by Eq. (7.1).

7.2.3.1 The Intensity Projection

Since the Fourier transform is distance preserving, minimising the distance in real space is the same as minimising the distance in Fourier space. Denote the positions of the Q grid points in Fourier space by $\mathbf{u}_q$. The set of Fourier samples $\{F_j(\mathbf{u}_q)\}$ for all $j = 1, \cdots, J$ and $q = 1, \cdots, Q$ can be considered as a point in a (complex) vector space of dimension JQ. The distance r between two such points $F_j(\mathbf{u}_{q0})$ and $F_j(\mathbf{u}_q)$ is given by

$$r^2 = \sum_{q=1}^{Q} \left(\sum_{j=1}^{J} ||F_j(\mathbf{u}_{q0}) - F_j(\mathbf{u}_q)||^2 \right) = \sum_{q=1}^{Q} d^2(\mathbf{u}_{q0}, \mathbf{u}_q) \tag{7.24}$$

where

$$d^2(\mathbf{u}_{q0}, \mathbf{u}_q) = \sum_{j=1}^{J} ||F_j(\mathbf{u}_{q0}) - F_j(\mathbf{u}_q)||^2. \tag{7.25}$$

Therefore, minimising the distance r corresponds to separately minimising $d^2(\mathbf{u}_{q0}, \mathbf{u}_q)$ for each q in each of the Q (complex) subspaces of dimension J. For each q the Fourier intensity data constraint is

$$I(\mathbf{u}_q) = \sum_{j=1}^{J} |F_j(\mathbf{u}_q)|^2, \tag{7.26}$$

which is a $2J$-dimensional hypersphere of radius $\sqrt{I(\mathbf{u}_q)}$ in a real $2J$-dimensional vector space, where the factor of two arises from taking into account the real and imaginary parts of $F_j(\mathbf{u}_q)$ for each j. The projection of a point in this space corresponds therefore to moving radially onto the hypersphere, i.e. the phase of each complex number $F_j(\mathbf{u}_q)$ is left unchanged but its amplitude is scaled by $\sqrt{I(\mathbf{u}_q)}/\sqrt{I(\mathbf{u}_{q0})}$, where $I(\mathbf{u}_{q0})$ is the Fourier intensity of the initial point.

In summary, the Fourier space projection utilised in the iterative phase retrieval algorithm for multiple independent images is achieved by modifying the Fourier transform at the ith iterate of each object, $F_j^{(i)}(\mathbf{u})$, by

$$F_j^{(i+1)}(\mathbf{u}) = \sqrt{\frac{I(\mathbf{u})}{\sum_{j=1}^{J} |F_j^{(i)}(\mathbf{u})|^2}} F_j^{(i)}(\mathbf{u}), \tag{7.27}$$

which collapses to the usual Fourier space projection for phase retrieval of a single object,

$$F_1^{(i+1)}(\mathbf{u}) = \sqrt{\frac{I(\mathbf{u})}{|F_1^{(i)}(\mathbf{u})|^2}} F_1^{(i)}(\mathbf{u}), \tag{7.28}$$

when $J = 1$. Inverse Fourier transforming after the above operation gives the Fourier space projection P_M.

7.2.4 Simulation Results

Simulations were conducted to illustrate the uniqueness results and the reconstruction algorithm described in the previous sections. Two- and three-dimensional objects with different supports were generated and their averaged Fourier intensity $I(\mathbf{u})$ calculated and used as data. Noise was not added to the data since the objective here is to investigate uniqueness and determination of the solution under ideal conditions. Since the objects may have different supports, the sampling grids for the objects need to be considered. As a matter of convenience, the sampling grid for each object is chosen here to be the same. In order to apply support constraints, and to satisfy the sampling requirements for phase problems, each object is embedded in a computational grid of the same size, that has a number of samples in each dimension that is at least twice the number of samples of the largest object support in that dimension. In the simulations described here, a factor of three times the number of samples of the largest object support in each dimension was used. The Fourier transform of each object is

calculated using the DFT, and the sampling grid for the Fourier transform of each object is thence the same. The known support of each object along with their averaged Fourier intensity was applied as constraints as outlined in the previous section. Positive real objects were used and a positivity constraint was sometimes applied as described in the results below. The DM algorithm was run for 10^4 iterations with $\beta = 0.8$, $\gamma_A = -1/\beta$ and $\gamma_B = 1/\beta$. Finally, 10 trials was carried out for each reconstruction with different random starting objects and the correct support.

Two error metrics were calculated to monitor convergence of the algorithm and the quality of the solution. The first metric monitors convergence of the algorithm and is the RMS difference between the intensity data and the averaged Fourier intensity of the reconstructed objects. This error metric is defined by

$$E = \sqrt{\frac{\sum_{\mathbf{u}} \left(I^{1/2}(\mathbf{u}) - \hat{I}^{1/2}(\mathbf{u}) \right)^2}{\sum_{\mathbf{u}} I(\mathbf{u})}} \tag{7.29}$$

where $I(\mathbf{u})$ is the intensity data and $\hat{I}(\mathbf{u})$ is the averaged Fourier intensity of the reconstructed objects, calculated at iteration i. This metric is small if the algorithm converges. The second error metric is the RMS difference between the set of reconstructed objects $\{\hat{f}_j(\mathbf{x})\}$ and the set of true objects $\{f_j(\mathbf{x})\}$, given by

$$e = \sqrt{\frac{\sum_j \sum_{\mathbf{x}} \left(f_j(\mathbf{x}) - \hat{f}_j(\mathbf{x}) \right)^2}{\sum_j \sum_{\mathbf{x}} f_j^2(\mathbf{x})}}. \tag{7.30}$$

This metric measures the accuracy of the reconstruction and converges to a small value if the correct solution is found. Note that E can be calculated in practice, but e cannot. However, e can be calculated in simulations where the true objects are known.

The final solution for each trial is obtained by taking the iterate where the error metric E is the smallest out of the 10^4 iterations. Converged runs (small E) that give a small e indicate a unique solution. Runs that give a small E and a large e indicate a non-unique solution. Because of their different characteristics, results are presented separately for the cases of identical and non-identical object supports.

To distinguish between objects that have not converged and objects that have converged but are linear combinations of the true objects $f_j(\mathbf{x})$ and their inversions in the origin $f_j(-\mathbf{x})$, the process described below is carried out. A linear least-squares fit on all samples of the true objects with the samples of the reconstructed objects and their centrosymmetric inversions is sought. The samples of the jth true object are first stacked into a vector $\mathbf{f}_j(\mathbf{x}) = [f_j(\mathbf{x}_1), \cdots, f_j(\mathbf{x}_Q)]^T$ and similarly for the samples of the jth reconstructed object, $\hat{\mathbf{f}}_j(\mathbf{x}) = [\hat{f}_j(\mathbf{x}_1), \cdots, \hat{f}_j(\mathbf{x}_Q)]^T$ and its inverted version $\hat{\mathbf{f}}_j(-\mathbf{x}) = [\hat{f}_j(-\mathbf{x}_1), \cdots, \hat{f}_j(-\mathbf{x}_Q)]^T$. Writing the coefficients of the jth row of the ma-

trices A and B developed in Sec. 7.2.2 as a vector $\mathbf{c} = [a_{j1}, \cdots, a_{jJ}, b_{j1}, \cdots, b_{jJ}]^T$ and forming the matrix $\mathrm{M} = [\, \hat{\mathbf{f}}_1(\mathbf{x}) \,|\, \cdots \,|\, \hat{\mathbf{f}}_J(\mathbf{x}) \,|\, \hat{\mathbf{f}}_1(-\mathbf{x}) \,|\, \cdots \,|\, \hat{\mathbf{f}}_J(-\mathbf{x}) \,]$, gives

$$\mathbf{f}_j(\mathbf{x}) = \mathrm{M}\mathbf{c}, \tag{7.31}$$

which expresses the samples of the true objects $\mathbf{f}_j(\mathbf{x})$ in terms of linear combinations of the samples of the reconstructed objects and their inversions, $\hat{\mathbf{f}}_j(\mathbf{x})$ and $\hat{\mathbf{f}}_j(-\mathbf{x})$.

The normal equation that solves for the coefficients in a linear least-squares problem such as that of Eq. (7.31) gives

$$\mathbf{c} = (\mathrm{M}^T\mathrm{M})^{-1}\mathrm{M}^T\mathbf{f}_j(\mathbf{x}), \tag{7.32}$$

which is the jth row of matrices A and B as mentioned above. Performing the operation for all $j = 1, \cdots, J$ therefore yields the matrices A and B by linear least-squares fitting which can then be used to "correct for" the effect of the reconstructed objects being linear combinations of the original objects and their inverted versions by application of Eqs. (7.22) and (7.23), where the quantity $\mathbf{g}(\mathbf{x})$ in those equations denote reconstructed objects that have converged. The operation described here is referred to below as "linear-combination correction" and only applies to objects that have the same support. Reconstructions before the application of linear-combination correction are referred to as "raw" reconstructions.

7.2.4.1 Identical Object Supports

The first simulation is for the two square 128×128 sample objects shown in Fig. 7.3(a). The reconstruction algorithm was run as described in the previous section, both with and without a positivity constraint applied. The algorithm converged in all cases ($E \sim 1 \times 10^{-3}$) but in no cases were good reconstructions obtained ($e \sim 5 \times 10^{-1}$). The real space error decreased after linear least squares fitting ($e \sim 4 \times 10^{-1}$), but the reconstruction still clearly does not resemble the original object. Typical reconstruction with and without positivity applied is shown in Fig. 7.3(b) and (c) respectively. This result confirms the expectation of non-uniqueness since this case is marginal with $\Omega = 1$, even with positivity.

The second simulation is for the case of two objects with identical centrosymmetric supports with $\Omega > 1$. The supports are a square of 128×128 samples with a central square of 64×64 samples excluded (a square donut) as shown in Fig. 7.4(a), which gives $\Omega = 1.33$. The algorithm was run both with and without positivity applied. With the positivity constraint, the algorithm converged in 9/10 of the trials ($E \sim 2 \times 10^{-7}$) and all of these gave reconstructions that can be seen to be mixtures of both images ($e \sim 1 \times 10^{-1}$). After correcting for the linear combinations, all of these were shown to be good solutions with $e \sim 3 \times 10^{-6}$. An example reconstruction is shown in

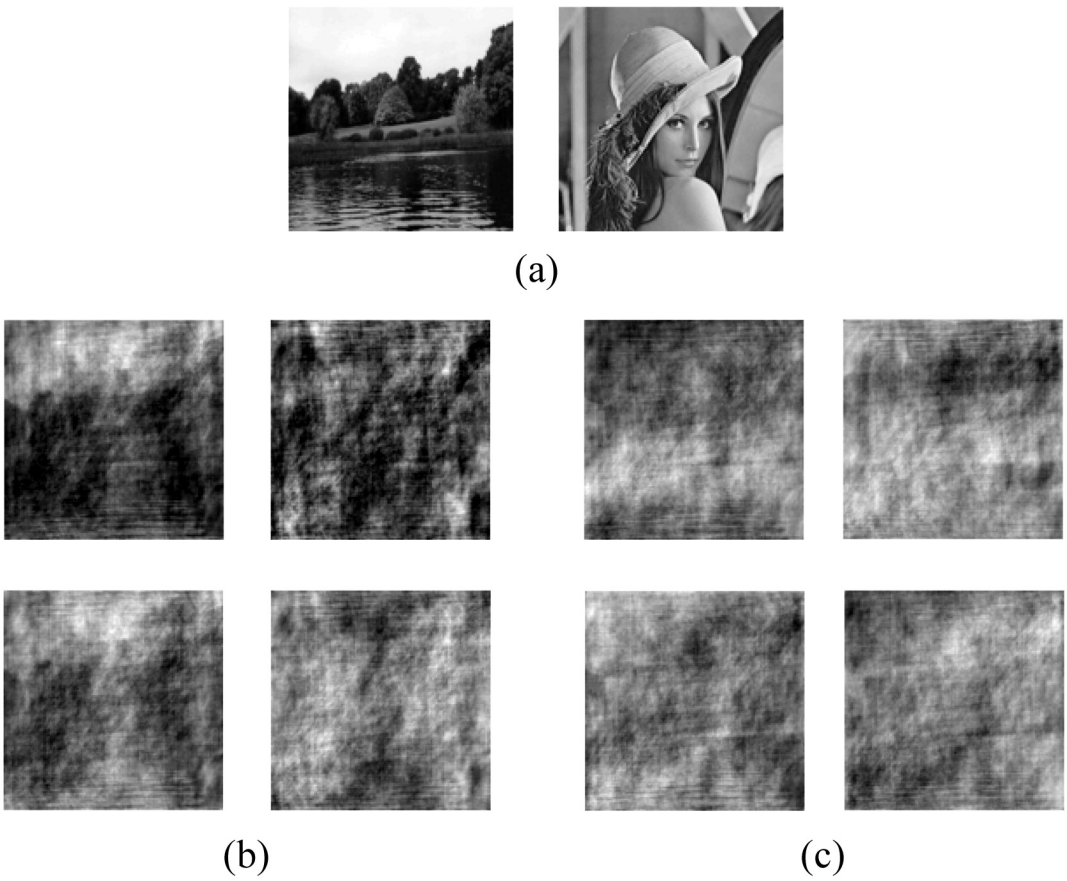

Figure 7.3 Reconstruction results for two square objects with identical supports. (a) Original objects, (b) example raw reconstructions (top row) and linear-combination corrected results (bottom row) with positivity applied and (c) example raw reconstructions (top row) and corrected results (bottom row) without positivity applied.

Fig. 7.4(b). The weak contamination of each reconstructed object by the other object and its inversion about the origin is consistent with the analysis given in Sec. 7.2.2 for objects with a centrosymmetric support. Without the positivity constraint, the algorithm converged in 4/10 trials ($E \sim 1 \times 10^{-8}$) with negative pixel values in the solution. Inverting the values of the pixels shows significant mixing has occurred with a real space error of $e \sim 3 \times 10^{-1}$. After correcting for the linear combinations, all of these were shown to be good solutions ($e \sim 2 \times 10^{-7}$).

The third simulation is for two right angle triangular objects with the two sides that are perpendicular to each other 128 samples long each, as shown in Fig. 7.5(a). The algorithm was run with and without positivity applied. With the positivity constraint, the algorithm converged in 8/10 of the trials ($E \sim 3 \times 10^{-6}$) and in all of these the correct objects were reconstructed ($e \sim 3 \times 10^{-2}$) with the real space error further reduced after correcting for linear combinations ($e \sim 3 \times 10^{-5}$). Example reconstructions are shown in Fig. 7.5(b) and (c). This result is consistent with the theory described before since in this case $\Omega = 1.5$, and without centrosymmetric supports, the

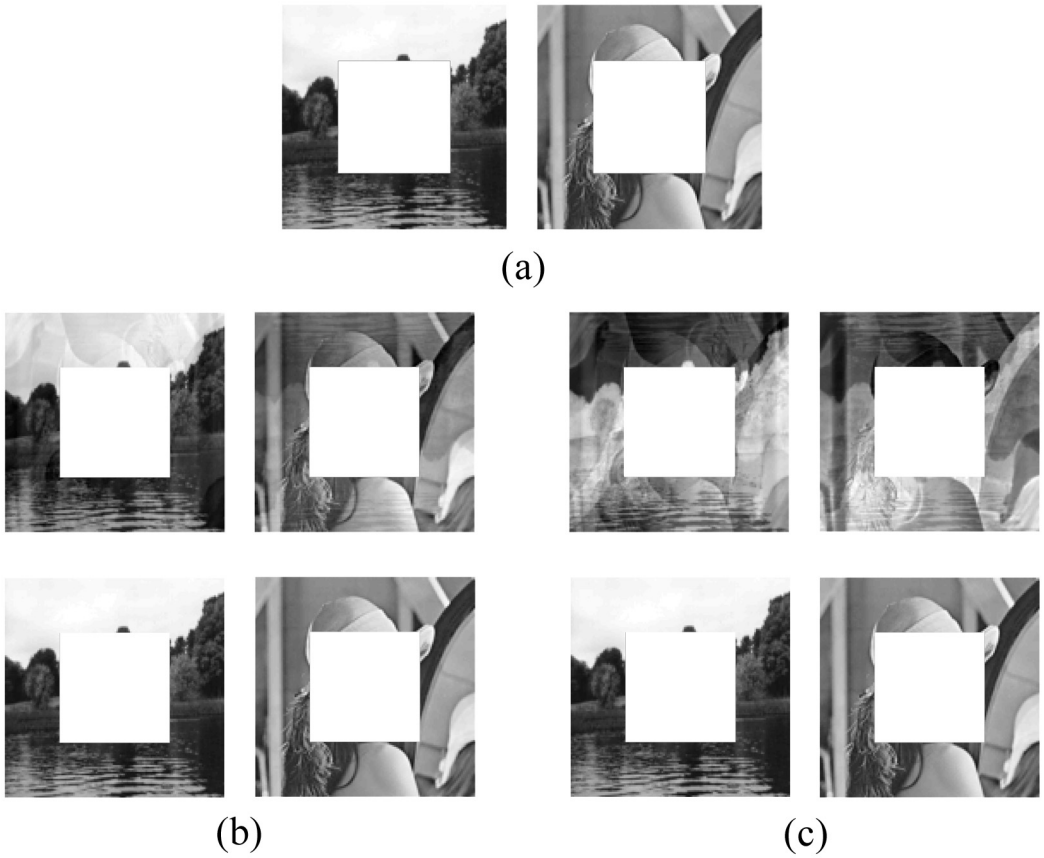

Figure 7.4 Reconstruction results for two square donut objects with identical supports. (a) Original objects, (b) example raw reconstructions (top row) and linear-combination corrected results (bottom row) with positivity applied and (c) example raw reconstructions (top row) and corrected results (bottom row) without positivity applied.

only mixing that can occur is between the two upright images. Without the positivity constraint, the algorithm converged in 8/10 of the trials ($E \sim 4 \times 10^{-6}$) and all of these gave reconstructions that have inverted pixels values, as can be seen in Fig. 7.5(c). Good real space errors after inverting the pixel values were attained ($e \sim 8 \times 10^{-2}$), and after linear combination correction, extremely good real space errors were obtained ($e \sim 3 \times 10^{-5}$).

The fourth set of simulations is for 3-D objects with $16 \times 16 \times 16$ samples each. The objects were generated by taking sections from a downsampled volume of the AQP1 protein electron density map [Ren et al. 2000] that was used in Sec. 5.5. Simulations were run for sets of 2, 3 and 4 objects. A resultant reconstruction for two objects is shown in Fig. 7.6. The algorithm converged for 10/10 trials ($E \sim 3 \times 10^{-4}$). Fairly poor raw reconstructions were obtained in all cases ($e \sim 2 \times 10^{-1}$) however analysis of the reconstructions shows that they are a linear combination of the objects and their inversions about the origin. Once corrected for, the real space error decreased to $e \sim 2 \times 10^{-3}$.

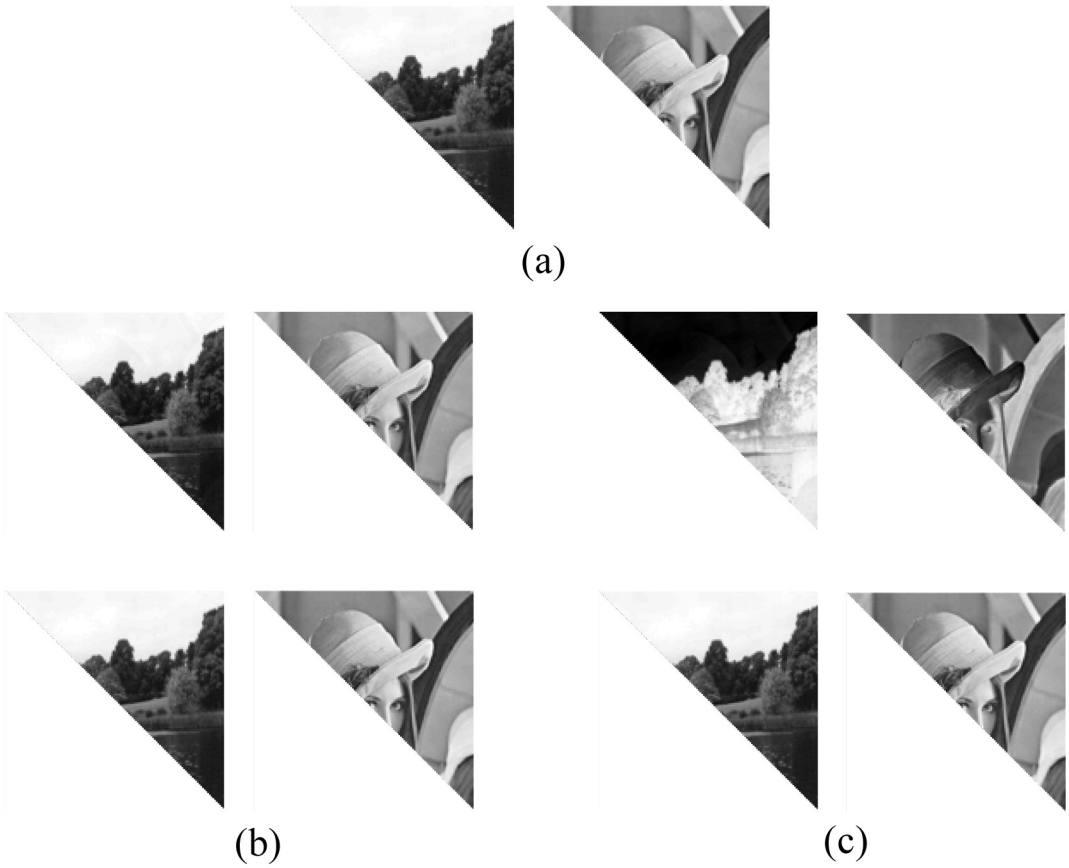

Figure 7.5 Reconstruction results for two triangular objects with identical supports. (a) Original objects, (b) example raw reconstructions (top row) and linear-combination corrected results (bottom row) with positivity applied and (c) example raw reconstructions (top row) and corrected results (bottom row) without positivity applied.

The results for three objects are shown in Fig. 7.7. The algorithm converged for 4/10 trials ($E \sim 3 \times 10^{-4}$) and again, poor reconstructions were obtained in all cases initially ($e \sim 3 \times 10^{-1}$) but after the least-squares fitting to find the linear combinations, these indeed correspond to a linear combination of the correct solutions ($e \sim 8 \times 10^{-3}$). For 4 objects, the algorithm converged in 10/10 trials ($E \sim 3 \times 10^{-6}$) but no correct solutions were obtained ($e \sim 4 \times 10^{-1}$) even after linear-combination correction ($e \sim 3 \times 10^{-1}$), which is expected since this case is marginal with $\Omega = 1$.

As described in Sec. 7.2.2.1, positivity is expected to exhibit better constraining on the solution if the objects have a wide range of sample values. This property is particularly significant for objects with centrosymmetric supports and for more objects, where there is more opportunity for the linear mixing of the objects, and their inverted versions, to occur in the solution. This point was investigated by repeating the reconstructions for the case of 3-D objects described above, for objects that have sample values chosen independently from a uniform distribution. The same simulations were run as described, and the results for two and three objects are shown in Figs. 7.8 and

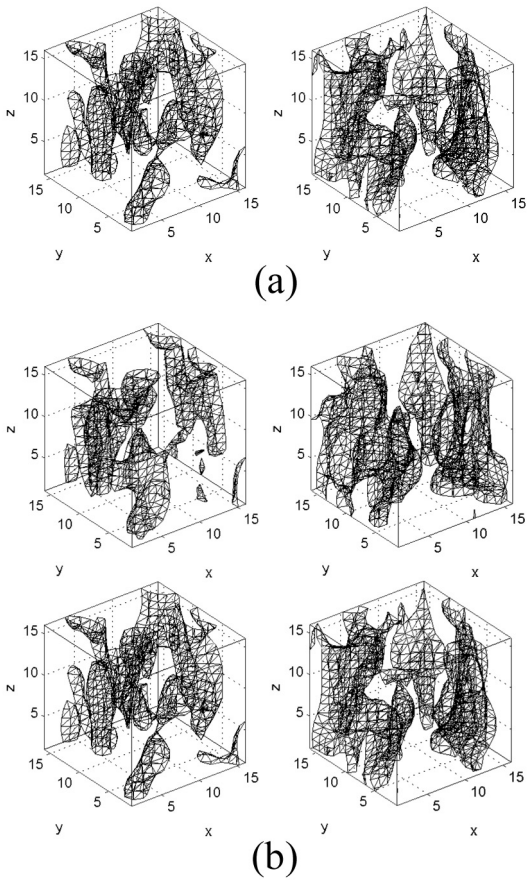

Figure 7.6 Reconstruction results for two 3-D objects with identical supports. (a) Original objects and (b) example raw reconstructions (top row) and linear-combination corrected results (bottom row) with positivity applied.

7.9, respectively. For two objects, 10/10 runs converged ($E \sim 5 \times 10^{-9}$) and good raw reconstructions ($e \sim 4 \times 10^{-4}$) were obtained, and for three objects, 10/10 runs converged ($E \sim 8 \times 10^{-7}$) which were all good raw reconstructions ($e \sim 1 \times 10^{-3}$). These results demonstrate the stronger constraining power of positivity for objects with a wide variety of sample values.

The final set of simulations in this section is used to illustrate the implications of identical, noncentrosymmetric supports in 3-D. The object support consists of a $12 \times 12 \times 12$ sample cube with an $8 \times 8 \times 6$ sample cuboid on one of its face as can be seen in Figs. 7.10(a) and 7.11(a). Simulations were run for two and three such objects in the same way as for the 3-D simulations above. The constraint ratios are $\Omega = 2.2$ and $\Omega = 1.47$, for the two and three object cases, respectively. These values for the continuous constraint ratio was obtained by plotting Ω_{D} for increasing scalings of the noncentrosymmetric object and extrapolating the curve to an infinite scale as was done in Sec. 7.2.1.2. The extrapolation gives Ω slightly larger than 4.4 for one copy of the single noncentrosymmetric object support in question, hence yielding the values for the two and three objects cases as mentioned above. For two objects, the

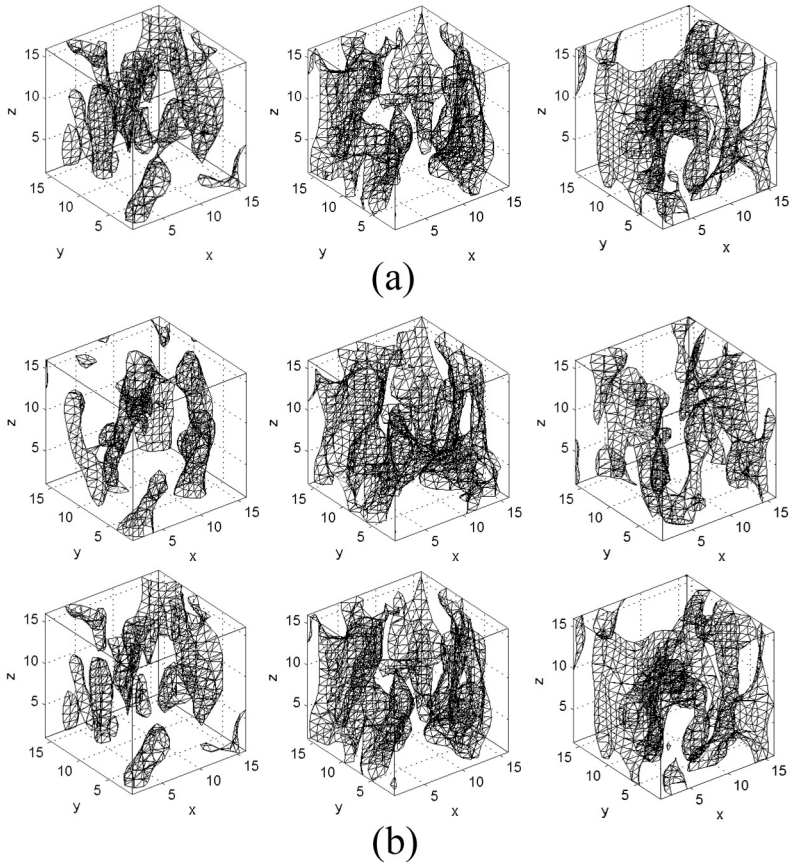

Figure 7.7 Reconstruction results for three 3-D objects with identical supports. (a) Original objects and (b) example raw reconstructions (top row) and linear-combination corrected results (bottom row) with positivity applied.

algorithm converged in 10/10 runs ($E \sim 2 \times 10^{-4}$) and all of them were good raw reconstructions ($e \sim 4 \times 10^{-3}$). For three objects, the algorithm converged in 10/10 runs ($E \sim 1 \times 10^{-4}$) and again all gave good raw reconstructions ($e \sim 7 \times 10^{-2}$). Comparison with the cases of centrosymmetric supports described previously shows that significantly better likelihood of good raw reconstructions were obtained. This result is as expected due to the exclusion of linear combinations of objects inverted about the origin as a result of the noncentrosymmetric support.

In summary, the results presented here support the theory for the case of identical object supports described in Sec. 7.2.2. Unique raw reconstructions are obtained if $\Omega > 1$, the support is noncentrosymmetric, and a positivity constraint is applied. If the support is centrosymmetric, the solution is less constrained, but raw reconstructions are close to the original objects. The positivity constraint is more effective for objects exhibiting a wide range of sample values, and a unique solution is obtained in such cases. Unique solutions are unlikely in the absence of positivity, even with a favourable constraint ratio ($\Omega > 1$). Good solutions are not obtained if $\Omega = 1$.

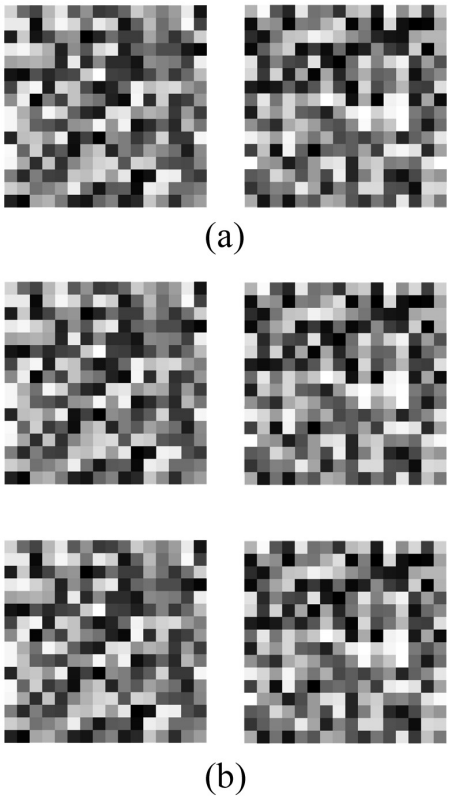

Figure 7.8 Reconstruction results for two 3-D objects with identical supports and uniformly distributed pixel values. $z = 1$ slice of (a) the original objects, (b) example raw reconstructions with positivity applied (top row) and the same reconstructions corrected for linear combinations (bottom row).

7.2.4.2 Different Object Supports

Consider now the case of objects with different supports. The first two simulations are for the case of two 2-D objects corresponding to cases A and B described in Sec. 7.2.1.3. The first simulation is for two objects with supports of 128×128 and 60×60 samples as shown in Fig. 7.12(a), which corresponds to case A in Sec. 7.2.1.3, and calculating the constraint ratio gives $\Omega = 1.64$. With positivity, the algorithm converged in 10/10 trials ($E \sim 1 \times 10^{-4}$) and good reconstructions were obtained in each case ($e \sim 4 \times 10^{-3}$). An example reconstruction is shown in Fig. 7.12(b). For the case without positivity, the algorithm converged in 4/10 trials ($E \sim 1 \times 10^{-7}$) and good reconstructions with inverted pixel values were obtained ($e \sim 5 \times 10^{-6}$). An example reconstruction is shown in Fig. 7.12(c).

The second simulation is for two objects with supports of 128×60 and 60×128 samples as shown in Fig. 7.13(a), which corresponds to case B in Sec. 7.2.1.3, and calculating the constraint ratio gives $\Omega = 1.44$. With positivity, the algorithm converged in 2/10 trials ($E \sim 3 \times 10^{-7}$) and good reconstructions were obtained in each case ($e \sim 7 \times 10^{-6}$). An example reconstruction is shown in Fig. 7.13(b). For

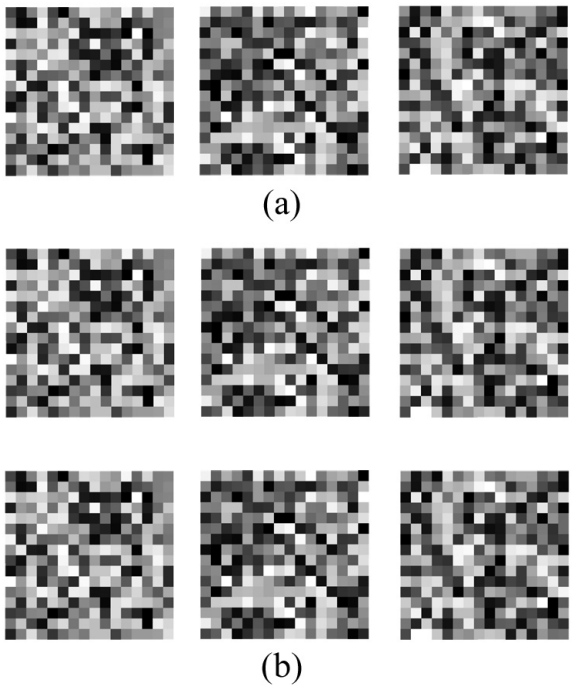

Figure 7.9 Reconstruction results for three 3-D objects with identical supports and uniformly distributed pixel values. $z = 1$ slice of (a) the original objects, (b) example raw reconstructions with positivity applied (top row) and the same reconstructions corrected for linear combinations (bottom row).

the case without positivity, the algorithm converged in 4/10 trials ($E \sim 2 \times 10^{-3}$) and good reconstructions with inverted pixel values were obtained in all cases. An example reconstruction is shown in Fig. 7.13(c). In this case then, good reconstructions were obtained both with and without positivity. These results together are consistent with the theory described previously since $\Omega > 1$.

The third simulation was conducted with three objects with dimensions 32×32, 16×64 and 64×16 pixels as shown in Fig. 7.14(a). Calculation of the constraint ratio as described in Sec. 7.2.1.3 gives $\Omega = 1.33$. Applying positivity, convergence was obtained in 2/10 trials ($E \sim 5 \times 10^{-4}$) and good reconstructions were obtained in all cases ($e \sim 9 \times 10^{-3}$). Example reconstructions are shown in Fig. 7.14(b). These results are as expected since again $\Omega > 1$.

Consider now the case of 3-D objects with different supports. Two sets of simulations were conducted, one with two objects of dimensions $20 \times 20 \times 20$ samples and $12 \times 12 \times 12$, and the other with three objects of dimensions $20 \times 20 \times 20$, $16 \times 16 \times 16$ and $12 \times 12 \times 12$ samples, as shown in Figs. 7.15(a) and 7.16(a), respectively. The constraint ratio is $\Omega = 3.29$ for the first case and $\Omega = 2.31$ for the second case. For the case of two objects, the algorithm converged in all 10 trials ($E \sim 4 \times 10^{-4}$) and good reconstructions were obtained in all cases ($e \sim 2 \times 10^{-3}$). An example reconstruction is shown in Fig. 7.15(b). For the case of three objects, the algorithm converged in all

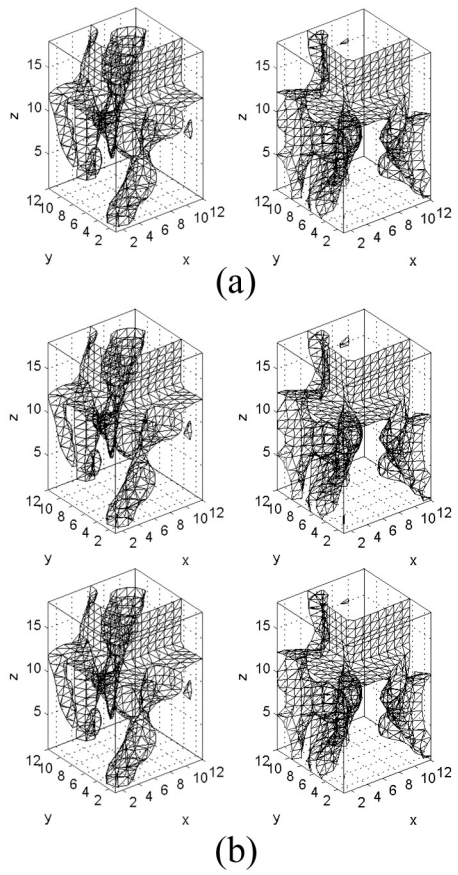

Figure 7.10 Reconstruction results for two 3-D objects with identical non-centrosymmetric supports as described in the text. (a) Original objects and (b) example raw reconstructions (top row) and linear combination corrected results (bottom row) with positivity applied.

10 trials as well ($E \sim 3 \times 10^{-4}$) and good reconstructions were obtained in all cases ($e \sim 3 \times 10^{-3}$). An example reconstruction is shown in Fig. 7.16(b). These results emphasise the uniqueness of the case with different supports, despite all the supports being centrosymmetric.

In principle, $\Omega > 1$ is strictly sufficient for a unique solution to the phase problem. However, in practice, even in the ideal case, the algorithm will converge to a near solution if Ω is close to unity. A practical question is then, how close to unity can Ω be for unique phase retrieval in the ideal case with no noise? Since in the cases studied here, Ω can be controlled by varying the shapes of the object supports, the relationship between uniqueness and Ω can be studied. This was done as follows. Two objects were reconstructed from their averaged Fourier intensity, one of which is square with a fixed size of 32×32 samples. The other object is rectangular with a size that varies between 32×32 samples and 64×16 samples in such a way that the number of samples are made to stay approximately the same (by rounding one side to the nearest integer after dividing 32^2 by the length of the other constantly varying side). In this way, the value of Ω for the reconstruction problem varies between 1 and 1.5. For each

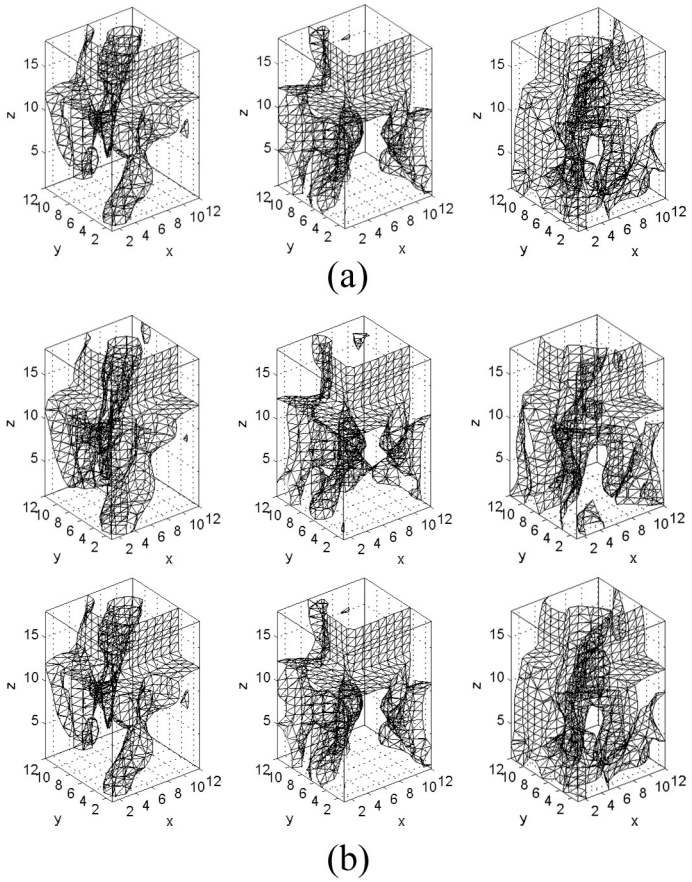

Figure 7.11 Reconstruction results for three 3-D objects with identical non-centrosymmetric supports as described in the text. (a) Original objects and (b) example raw reconstructions (top row) and linear combination corrected results (bottom row) with positivity applied.

integer value of the short side of the rectangular object, 100 trials of the algorithm were run, each with 10^4 iterations, and with a positivity constraint applied. For each trial, the runs with good convergence were selected and the mean reconstruction error was calculated. The reconstruction errors E and e are plotted in Fig. 7.17 versus the value of Ω_{D} for that trial. Inspection of the figure shows poor reconstructions for the values of the constraint ratio near unity where the two support shapes are the same, and the real space error e decreasing as Ω_{D} increases to about 1.2, after which good reconstructions are consistently obtained. The Fourier intensity error E that measures the convergence of the phase retrieval algorithm is low for large values of the constraint ratio, increases as the support shapes become identical and then decreases again once Ω_{D} drops to below 1. The explanation for the ease of convergence when the constraint ratio is less than one is that the problem becomes under-determined and thus many other solutions are able to satisfy the data. The results show the expected behaviour of the solution with Ω. In practice, the dependence on Ω is expected to be complicated and will likely depend on the nature and the size of the particular problem at hand, although Fig. 7.17 should show the typical kind of behaviour to be expected.

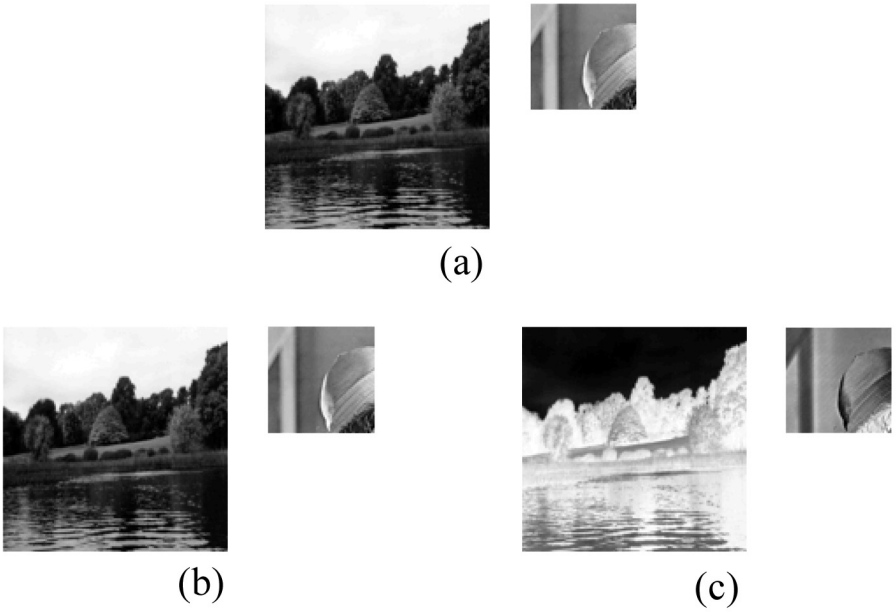

Figure 7.12 Reconstruction results for two square objects with different supports. (a) Original objects, (b) example reconstructions with positivity applied, and (c) reconstructions without positivity applied.

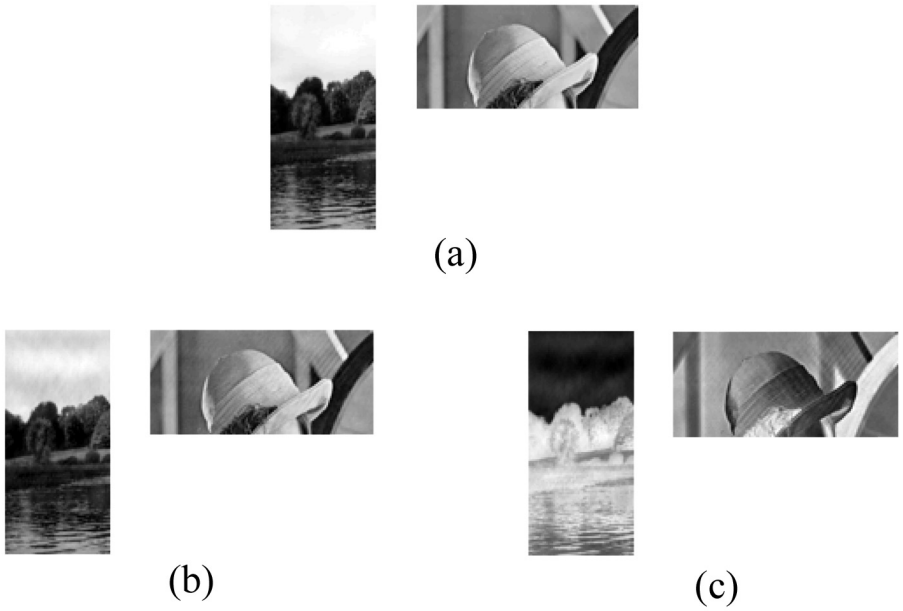

Figure 7.13 Reconstruction results for two rectangular objects with different supports. (a) Original objects, (b) example reconstructions with positivity applied, and (c) reconstructions without positivity applied.

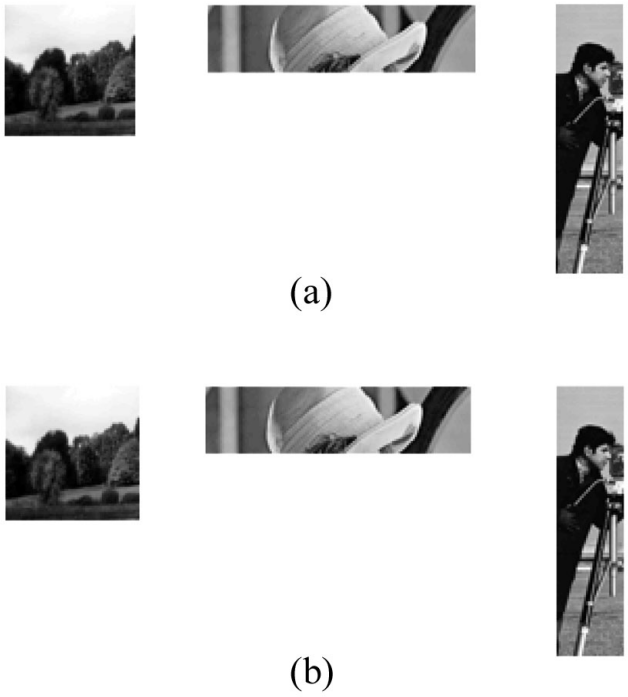

Figure 7.14 Reconstruction results for three 2-D objects with different supports. (a) Original objects and (b) example reconstructions with positivity applied.

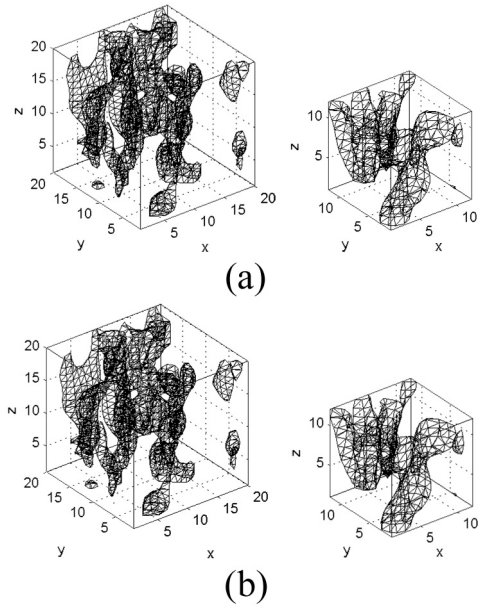

Figure 7.15 Reconstruction results for two 3-D objects with different centrosymmetric supports. (a) Original objects and (b) example reconstructions with positivity applied.

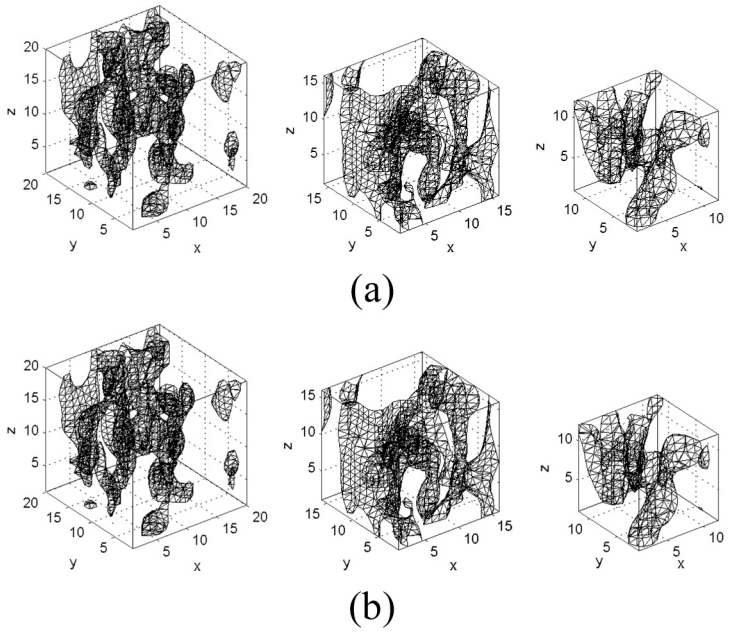

Figure 7.16 Reconstruction results for three 3-D objects with different centrosymmetric supports. (a) Original objects and (b) example reconstructions with positivity applied.

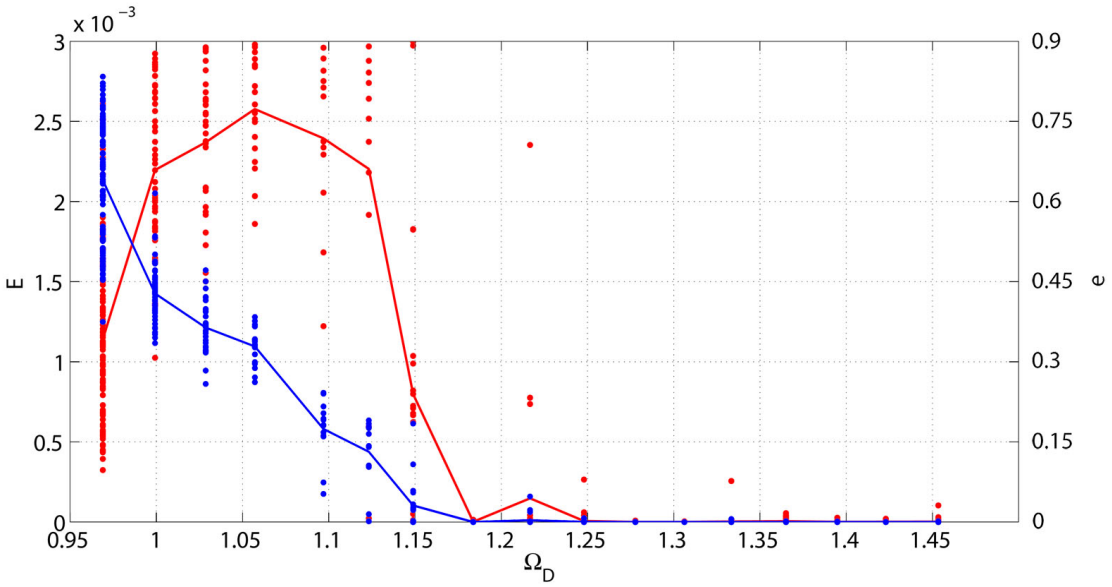

Figure 7.17 Reconstruction errors E (red) and e (blue) as a function of Ω_{D} for a 32×32 square object and a rectangular object of variable shape. The length of the sides of the rectangular shape is varied in such increments that the total number of samples in the rectangular shape stays approximately 32^2 until it becomes square. One hundred trials with different random starts were done for each increment and trials resulting in reconstructions with an E below the chosen threshold of $E = 3 \times 10^{-3}$ were kept. The mean errors from the trials that were kept are shown as solid lines and the points denote errors from the individual kept trials.

Table 7.1 Simulation results summary.

Fig.	Support shapes*	Dimen.	Identical supports	Centro-symm.	No. of objects	Ω	Ω_D	With Positivity				Without Positivity			
								Convg. rate	E	e	e_C†	Convg. rate	E	e	e_C†
7.3	Square	2	Y	Y	2	1	0.99	100%	9.17e-4	4.89e-1	3.95e-1	100%	5.05e-4	5.55e-1	4.88e-1
7.4	Sq. Donut	2	Y	Y	2	1.33	1.32	90%	2.10e-7	1.36e-1	3.04e-6	40%	1.21e-4	2.58e-1	1.54e-7
7.5	Triangle	2	Y	N	2	1.5	1.48	80%	3.30e-6	2.94e-2	3.09e-5	80%	3.67e-6	7.76e-2	3.10e-5
7.6	Cube	3	Y	Y	2	2	1.82	100%	2.52e-4	1.52e-1	1.54e-3	-	-	-	-
7.7	Cube	3	Y	Y	3	1.33	1.21	40%	2.67e-4	2.37e-1	7.55e-3	-	-	-	-
-	Cube	3	Y	Y	4	1	0.91	100%	2.58e-6	4.22e-1	3.01e-1	-	-	-	-
7.8	Cube (R)	3	Y	Y	2	2	1.82	100%	5.34e-9	3.92e-4	1.10e-8	-	-	-	-
7.9	Cube (R)	3	Y	Y	3	1.33	1.21	100%	7.96e-7	1.36e-3	2.94e-5	-	-	-	-
7.10	Noncent.	3	Y	N	2	2.20	1.95	100%	2.07e-4	4.04e-3	9.00e-4	-	-	-	-
7.11	Noncent.	3	Y	N	3	1.47	1.30	100%	1.49e-4	6.78e-2	2.03e-3	-	-	-	-
7.12	Square	2	N	Y	2	1.64	1.63	100%	1.27e-4	3.79e-3	NA	40%	1.82e-7	5.30e-6	NA
7.13	Rectangle	2	N	Y	2	1.53	1.51	20%	3.47e-7	7.00e-6	NA	40%	4.84e-7	1.34e-5	NA
7.14	Rectangle	2	N	Y	3	1.33	1.31	20%	4.65e-4	9.32e-3	NA	-	-	-	NA
7.15	Cube	3	N	Y	2	3.29	3.21	100%	4.31e-4	2.07e-3	NA	-	-	-	NA
7.16	Cube	3	N	Y	3	2.31	2.26	100%	3.29e-4	2.70e-3	NA	-	-	-	NA

* R indicates that the objects have random values

† linear-combination corrected errors.

NA denotes not applicable

- means not calculated.

Representative error plots for the phase retrieval simulations described above are displayed in Figs. 7.18 and 7.19. These figures show E and e for one particular DM algorithm run as a function of the iterative projection algorithm iteration. Each figure contains error metrics for two particular multiple object phase retrieval scenarios that differ only in a single feature, enabling comparison. The features differed are: random vs not random (Fig. 7.18) and 128×60 and 60×128 objects vs 128×128 and 60×60 objects (Fig. 7.19). Details of the other parameters that are kept the same in each phasing scenario are given in the captions. Note that the real space error in these plots (blue curves) are not linear-combination corrected (where applicable, i.e. Fig. 7.18), hence their values may seem somewhat large despite convergence having in fact been achieved as indicated by a low E.

Table. 7.17 on the previous page provides a summary of all the numerical experiments carried out in the above two subsections on reconstructing multiple independent objects from their summed Fourier intensity.

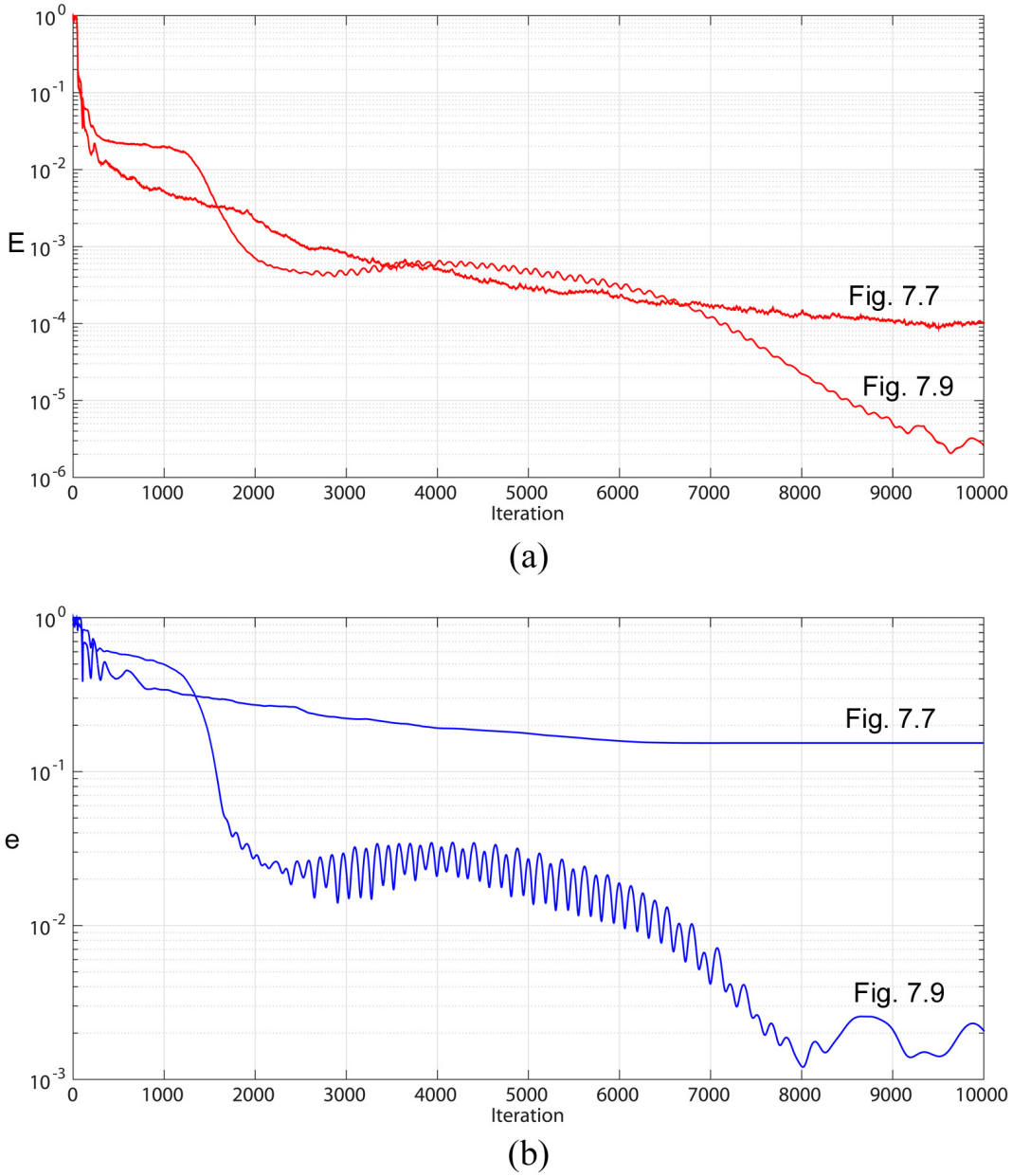

Figure 7.18 Reconstruction errors E (a) and e (b) for three 3-D objects with identical supports and positivity. The two curves correspond to the cases for random objects (Fig. 7.9), and non-random objects (Fig. 7.7).

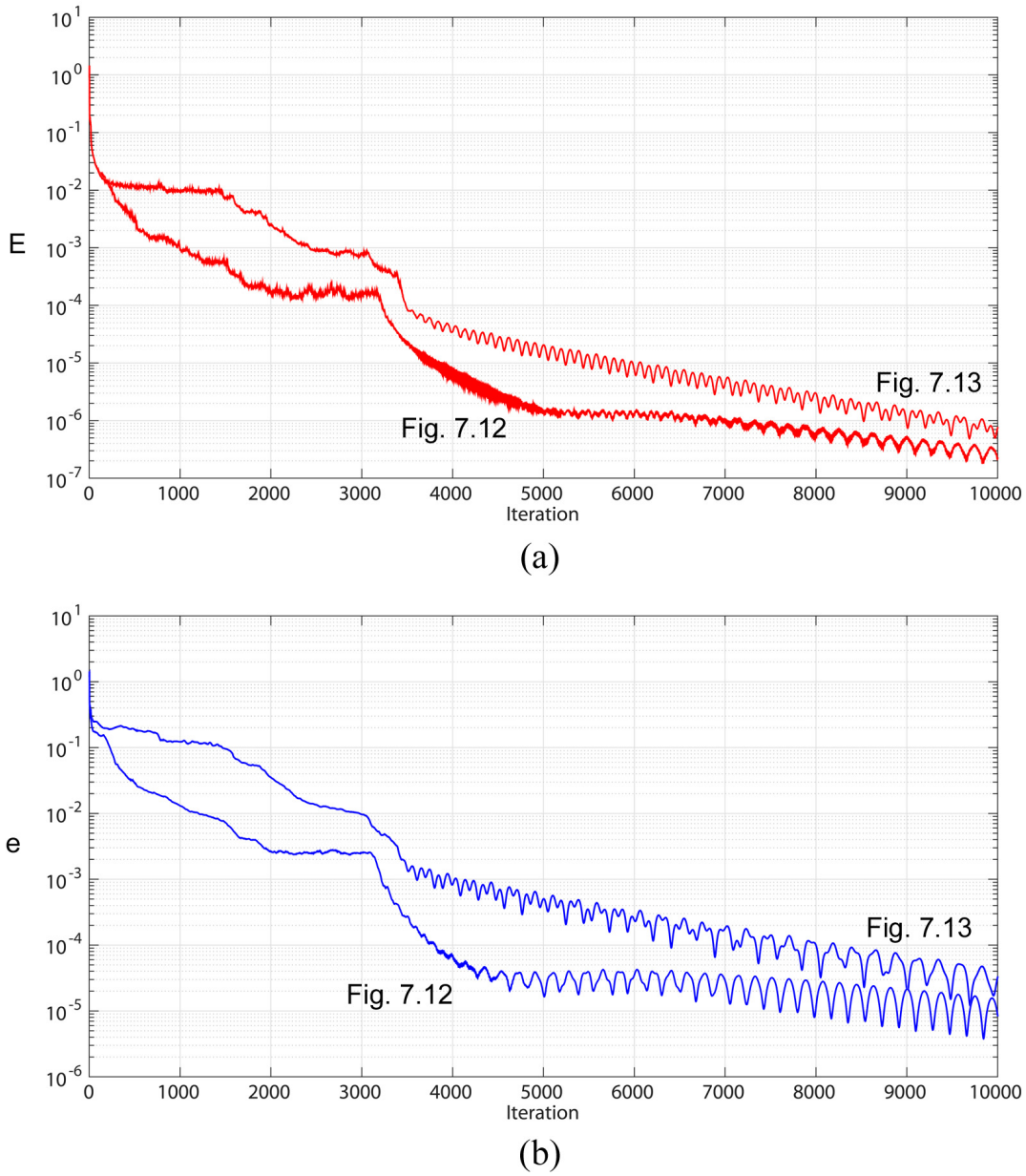

Figure 7.19 Reconstruction errors E (a) and e (b) for two 2-D objects with different supports and positivity. The two curves correspond to the cases for 128×60 and 60×128 objects (Fig. 7.13), and 128×128 and 60×60 objects (Fig. 7.12).

7.3 PHASE RETRIEVAL FOR MULTIPLE UNIT CELLS

It was shown in Chapter 6 that if there is more than one molecule in the unit cell, then different unit cells can be defined, for example, by translationally wrapping the molecule in one unit cell, or equivalently, shifting the boundary of a unit cell so that it encompasses the next whole molecule. In the simplest case which will be considered here, a molecule corresponds to the asymmetric unit. The molecules in the unit cell are related to each other by rigid body rotations and translations. Thus, the electron density of the jth kind of unit cell $f_j(\mathbf{x})$ can be written in terms of the electron density of the asymmetric unit $a(\mathbf{x})$ as

$$f_j(\mathbf{x}) = \sum_{k=0}^{K-1} a(\mathrm{R}^k\mathbf{x} + \mathbf{t}_{jk}), \tag{7.33}$$

where $j = 1, \cdots, J$ indexes the different kinds of complete unit cell and there are K asymmetric units in the unit cell in total. Note that the set of rotations is the same for all unit cells but the set of translations are different. Assuming that there is no surface structure, the electron density of the crystal built from the jth kind of unit cell,

$$g_j(\mathbf{x}) = \sum_{\mathbf{m}} f_j(\mathbf{x} - \mathrm{L}\mathbf{m}) \tag{7.34}$$

can then be expressed as

$$g_j(\mathbf{x}) = \sum_{\mathbf{m}} \sum_{k=0}^{K-1} a(\mathrm{R}^k\mathbf{x} + \mathbf{t}_{jk} - \mathrm{R}^k\mathrm{L}\mathbf{m}), \tag{7.35}$$

where R^k and $\mathbf{t}_{jk}$ are the rotations and translations of the asymmetric unit required to construct the jth kind of unit cell.

The phase retrieval problem at hand is then to reconstruct one of the unit cells, or equivalently one asymmetric unit, from the diffracted intensity averaged over the different unit cells. The problem is thus related to that of reconstructing multiple objects from their averaged diffracted intensity as described in Sec. 7.2, with the following additional characteristics: (1) The supports of the objects are identical, (2) there are symmetry relationships within each object, and (3) there are relationships between the objects themselves according to the redefinition of the unit cell as described in Chapter 6. These characteristics modify the nature of the problem. Some investigation into the phase retrieval problem for multiple objects with the above mentioned characteristics has been made but the results are incomplete. The progress made so far is reported in this section.

7.3.1 Diffraction Formulation

It is assumed, as described in the previous section, that the diffracted intensity averaged over the different unit cells can be extracted from the ensemble averaged SFX diffraction data. A candidate approach is the division method described in Chapter 5. The averaged intensity is given by

$$I(\mathbf{u}) = \sum_{j=1}^{J} |F_j(\mathbf{u})|^2, \tag{7.36}$$

where $F_j(\mathbf{u})$ is now the Fourier transform of the electron density $f_j(\mathbf{x})$ of the jth kind of unit cell. As described above, $f_j(\mathbf{x})$, and therefore $F_j(\mathbf{u})$, are related to each other due to the symmetry relationships within each $f_j(\mathbf{x})$ and the relationships between different $f_j(\mathbf{x})$ by the particular definition of the jth unit cell.

It is instructive to consider here a particular case. Consider a 2-D crystal with P4 symmetry in the unit cell, and four translationally wrapped unit cells as described in Sec. 6.3.1, so that $J = 4$, $K = 4$ and the rotation matrix R is

$$\mathrm{R} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}. \tag{7.37}$$

Taking the unit cell $f_1(\mathbf{x})$ as the reference unit cell, the translation vectors $\mathbf{t}_{jk}$ are

$$\begin{aligned}
\mathbf{t}_{10} &= w[0,0]^T & \mathbf{t}_{20} &= w[0,-1]^T \\
\mathbf{t}_{11} &= w[0,-1]^T & \mathbf{t}_{21} &= w[0,0]^T \\
\mathbf{t}_{12} &= w[1,-1]^T & \mathbf{t}_{22} &= w[1,0]^T \\
\mathbf{t}_{13} &= w[1,0]^T & \mathbf{t}_{23} &= w[1,-1]^T \\
\\
\mathbf{t}_{30} &= w[1,-1]^T & \mathbf{t}_{40} &= w[1,0]^T \\
\mathbf{t}_{31} &= w[1,0]^T & \mathbf{t}_{41} &= w[1,-1]^T \\
\mathbf{t}_{32} &= w[0,0]^T & \mathbf{t}_{42} &= w[0,-1]^T \\
\mathbf{t}_{33} &= w[0,-1]^T & \mathbf{t}_{43} &= w[0,0]^T
\end{aligned} \tag{7.38}$$

where w is the length of one side of the asymmetric unit.

Note that the more intuitive definition of the rotation matrix being a clockwise 90° rotation,

$$\mathrm{R} = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}, \tag{7.39}$$

and the translation vectors being

$$\begin{aligned}
\mathbf{t}_{10} &= w[0,0]^T & \mathbf{t}_{20} &= w[1,0]^T \\
\mathbf{t}_{11} &= w[1,0]^T & \mathbf{t}_{21} &= w[0,0]^T \\
\mathbf{t}_{12} &= w[1,1]^T & \mathbf{t}_{22} &= w[0,1]^T \\
\mathbf{t}_{13} &= w[0,1]^T & \mathbf{t}_{23} &= w[1,1]^T \\
\\
\mathbf{t}_{30} &= w[1,1]^T & \mathbf{t}_{40} &= w[0,1]^T \\
\mathbf{t}_{31} &= w[0,1]^T & \mathbf{t}_{41} &= w[1,1]^T \\
\mathbf{t}_{32} &= w[0,0]^T & \mathbf{t}_{42} &= w[1,0]^T \\
\mathbf{t}_{33} &= w[1,0]^T & \mathbf{t}_{43} &= w[0,0]^T
\end{aligned} \tag{7.40}$$

corresponding to the actual motion of the letter P molecule in Fig. 7.20, would require Eq. (7.33) to be in the form

$$f_j(\mathbf{x}) = \sum_{k=0}^{K-1} a\big(\mathrm{R}^{-k}(\mathbf{x} - \mathbf{t}_{jk})\big). \tag{7.41}$$

In general, there will be three length parameters in 3-D, w_1, w_2 and w_3, with a cuboidal asymmetric unit having a support given by $\mathrm{rect}\big(\mathrm{W}^{-1}(\mathbf{x} - \mathbf{t}_{jk})\big)$ where W is a square diagonal matrix with w_1, w_2 and w_3 on its diagonal.

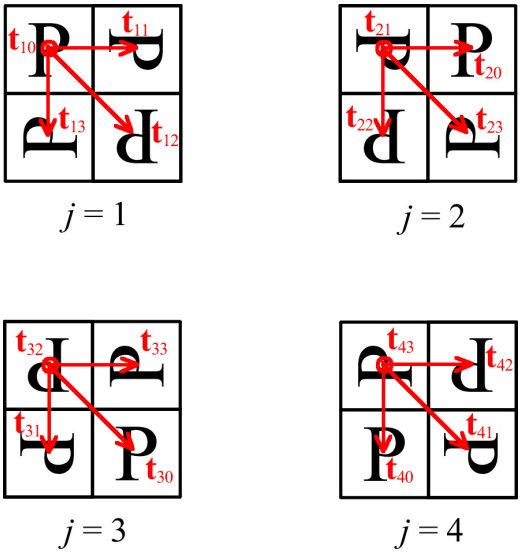

Figure 7.20 The four different kinds of unit cells and the "intuitive" translation vectors $\mathbf{t}_{jk}$ with respect to Eq. (7.41) for the P4 space group in 2-D. The origin of the coordinate system for which the $\mathbf{t}_{jk}$ is defined is at the centre of the top left asymmetric unit for each unit cell, and increases to the right (first component of the $\mathbf{t}_{jk}$ vector) and downwards (second component of the $\mathbf{t}_{jk}$ vector).

Note that for the plane space group P4, the four unit cells generated from translationally wrapping the asymmetric unit are all centrosymmetric. This implies that the Fourier transforms of these unit cells are real with phase values restricted to either 0 or π, rendering the phase problem easier to solve in general.

7.3.1.1 Reconstruction Algorithms

There are two possible approaches to the phase retrieval problem at hand, both of which can be adapted into the classical iterative phase retrieval scheme. The first follows directly from the independent image reconstruction problem described in Sec. 7.2, in which the individual unit cells (i.e. the objects) are reconstructed. An extra constraint in real space is required to enforce the symmetry relationships within and between the unit cells. A second approach has been proposed by Kirian et al. [2014] that involves directly reconstructing the single asymmetric unit $a(\mathbf{x})$, rather than the individual unit cells. The first approach is described in Sec. 7.3.1.2 and simulation results are presented in Sec. 7.3.1.3. Some analysis of the second approach is presented in Sec. 7.3.2.

7.3.1.2 Reconstruction Algorithm 1

For the first approach described above, the J different unit cells at the ith iteration are denoted $f_j^{(i)}(\mathbf{x})$ and their Fourier transforms $F_j^{(i)}(\mathbf{u})$. The ensemble averaged diffracted intensity at the ith iteration is then given by

$$I^{(i)}(\mathbf{u}) = \frac{1}{J}\sum_{j=1}^{J} |F_j^{(i)}(\mathbf{u})|^2 \ . \tag{7.42}$$

As described in the previous section, this is identical to the case of phase retrieval from independent images and the Fourier space projection P_M is given by the replacement

$$F_j^{(i+1)}(\mathbf{u}) = \sqrt{\frac{I(\mathbf{u})}{I^{(i)}(\mathbf{u})}} F_j^{(i)}(\mathbf{u}) \ . \tag{7.43}$$

The real space projection P_S for this approach has two steps: one, enforcing the support of the unit cells, and two, enforcing the symmetry constraints within and between the J unit cells by symmetry-averaging. The step of symmetry averaging involves finding all positions in the set of unit cells that must have the same electron density values due to space group symmetry and the relationships between the different unit cells. The average of the electron density values at those positions is taken and the value assigned back to those positions. These operations can be formalised as follows.

First, the set of unit cells at the start of the ith iteration of the phase retrieval algorithm, before being operated on by the real space projection P_S, will in general not have the correct symmetry. The unit cells can therefore be described by asymmetric

units with electron densities that vary over the different kinds of unit cell (indexed by j) and within individual unit cells (indexed by k). Denoting the electron density of these variable asymmetric units at the ith iteration of the phase retrieval algorithm by $a_{jk}^{(i)}(\mathbf{x})$, the electron density of the jth unit cell at the ith iteration is then given by Eq. (7.33) as

$$f_j^{(i)}(\mathbf{x}) = \sum_{k=0}^{K-1} a_{jk}^{(i)}(\mathrm{R}^k\mathbf{x} + \mathbf{t}_{jk}). \tag{7.44}$$

Un-doing the symmetry operation gives

$$a_{jk}^{(i)}(\mathbf{x}) = f_j^{(i)}\big(\mathrm{R}^{-k}(\mathbf{x} - \mathbf{t}_{jk})\big)\mathrm{rect}\big(\mathrm{W}^{-1}(\mathbf{x} - \mathbf{t}_{jk})\big). \tag{7.45}$$

Averaging over all K asymmetric units from all J unit cells thus yields the desired symmetry-averaged asymmetric unit, i.e.,

$$a^{(i+1)}(\mathbf{x}) = \frac{1}{JK}\sum_{j=1}^{J}\sum_{k=0}^{K-1} a_{jk}^{(i)}(\mathbf{x}). \tag{7.46}$$

Constructing all J symmetry-averaged unit cells using this asymmetric unit by Eq. (7.33), the electron density of the jth unit cell at the $(i+1)$th iteration of the phase retrieval algorithm after the projection P_S is hence

$$\begin{aligned} f_j^{(i+1)}(\mathbf{x}) &= \sum_{k=0}^{K-1} a^{(i+1)}(\mathrm{R}^k\mathbf{x} + \mathbf{t}_{jk}) \\ &= \frac{1}{JK}\sum_{k=0}^{K-1}\sum_{j_1=1}^{J}\sum_{k_1=0}^{K-1} a_{j_1k_1}^{(i)}(\mathrm{R}^k\mathbf{x} + \mathbf{t}_{jk}) \\ &= \frac{1}{JK}\sum_{k=0}^{K-1}\sum_{j_1=1}^{J}\sum_{k_1=0}^{K-1} f_{j_1}^{(i)}\big(\mathrm{R}^{-k_1}(\mathrm{R}^k\mathbf{x} + \mathbf{t}_{jk} - \mathbf{t}_{j_1k_1})\big)\mathrm{rect}\big(\mathrm{W}^{-1}(\mathrm{R}^k\mathbf{x} + \mathbf{t}_{jk} - \mathbf{t}_{j_1k_1})\big) \end{aligned} \tag{7.47}$$

by Eqs. (7.46) and (7.45). Equation (7.47) is the required symmetry-averaging step in the real space projection P_S.

7.3.1.3 Simulation Results

Simulations were conducted to investigate the phase retrieval algorithm described above. A 32×32 pixel Lena image was used as the asymmetric unit and the four different unit cell types were generated as shown in Fig. 7.20 and Fig. 6.13(b) from the previous chapter such that Lena represents the molecule. The Fourier intensity for the four unit cells was calculated using the discrete Fourier transform and averaged together to represent the data $I(\mathbf{u})$. Noise was not added since the objective here is to investigate uniqueness and the determination of solutions under ideal conditions.

The images were zero-padded by a factor of four in each direction to ensure that the Fourier intensities are sufficiently sampled. The known support, positivity, and the symmetry operations were applied as constraints in real space. The intensity data given by Eq. (7.36) is the constraint in Fourier space and was applied as described above. The DM algorithm with $\beta = 0.8$ was started such that the initial iterate has random pixel values in real space with the correct supports. The algorithm was run for 10^4 iterations and the error metrics E and e were calculated. The solution was obtained by taking the iterate where the metric E is a minimum.

The algorithm was run a number of times starting with different random images each time. The algorithm converged ($E \sim 10^{-6}$) for approximately 80% of the runs. For the converged runs, one of two different reconstructions was obtained, as shown in Fig. 7.21. One reconstruction (Fig. 7.21(a)) was visually indistinguishable from the true image. The other reconstruction (Fig. 7.21(b)) has a very different appearance. Analysis of the second reconstruction show that it is a linear combination of the first with the linear combination coefficient matrices

$$\mathrm{A} = \frac{1}{2}\begin{bmatrix} -1 & 1 & 1 & 1 \\ 1 & -1 & 1 & 1 \\ 1 & 1 & -1 & 1 \\ 1 & 1 & 1 & -1 \end{bmatrix} \tag{7.48}$$

and $\mathrm{B} = 0$. Out of those reconstructions that converged the original set of images was successfully reconstructed around half of the time and the linear-combination generated set the other half. The bottom row of Fig. 7.21 shows the averaged diffracted intensity for the two sets of unit cells which are visually identical with a normalised RMS error of around 10^{-7}. It appears that in this particular case the positivity constraint suppresses all other linear combinations and leaves only the two solutions for the P4 symmetry reconstruction investigated here.

For a set of J objects, the coefficient matrix A corresponding to the second solution has the form

$$\mathrm{A} = \frac{2}{J}\begin{bmatrix} 1-\frac{J}{2} & 1 & \cdots & 1 \\ 1 & 1-\frac{J}{2} & \cdots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \cdots & 1-\frac{J}{2} \end{bmatrix} = \frac{2}{J}[1]_{J,J} - \mathrm{I}, \tag{7.49}$$

where $[1]_{J,J}$ denotes the $J \times J$ all-ones matrix. It can be shown that A is indeed orthogonal. The second equality of Eq. (7.49) says in words that the alternate solution for the jth unit cell is equal to twice the average of the electron density from all J original unit cells minus the jth original unit cell. Note that the B matrix is 0 because in this case, all unit cells are equal to their own inversion about the origin, and so B is absorbed into A.

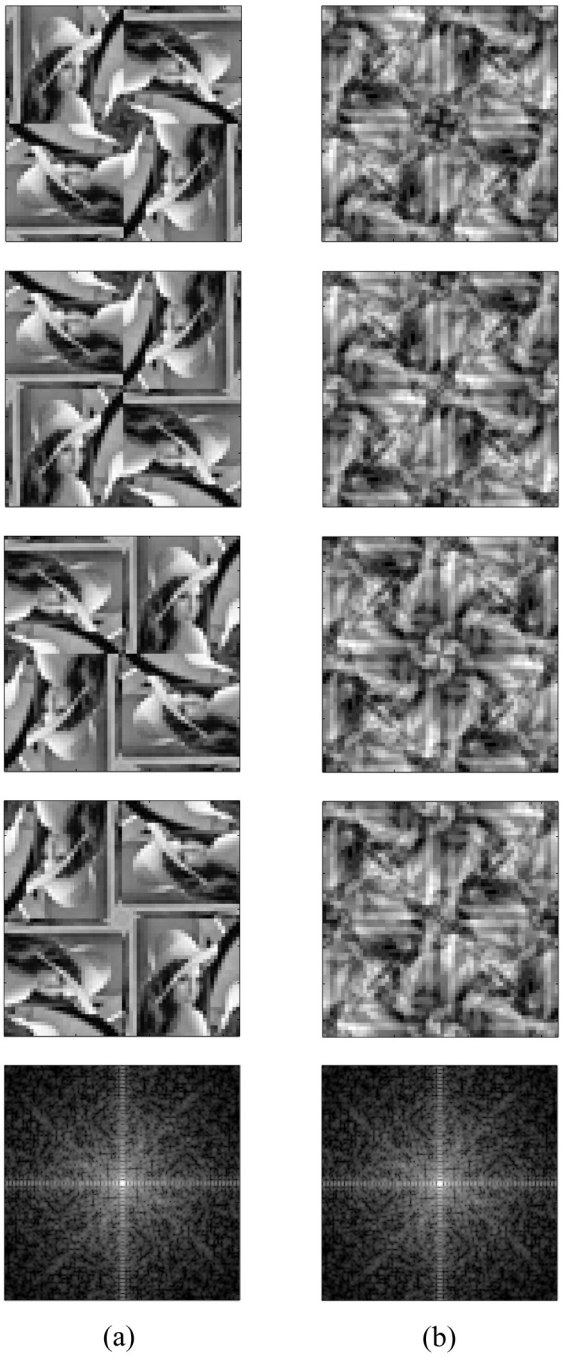

Figure 7.21 Reconstructions in the cases (a) where the algorithm converged on the original set of unit cells and (b) where the algorithm converged to a linear combination of the original unit cells as described in the text. The bottom row shows the averaged diffracted intensity for each reconstruction.

In summary, these results show that the reconstruction algorithm described is effective and that at least in this particular case there appears to be only a two-fold ambiguity in the solution. Clearly, more work is needed to assess the uniqueness properties of this problem further.

7.3.2 Reconstruction Algorithm 2

For the second approach, the electron density of the asymmetric unit $a(\mathbf{x})$ and its Fourier transform $A(\mathbf{u})$ are estimated at each iteration. This approach to phase retrieval reconstructs the asymmetric unit as opposed to the set of all unit cells as in the first algorithm described in Sec. 7.3.1.2. The advantage of this approach is that for K asymmetric units in a unit cell and J unit cells, the computational size of the problem is reduced by a factor of JK.

Referring to Eq. (7.35), the measured ensemble averaged diffraction intensity can be written in terms of $A(\mathbf{u})$ as

$$I(\mathbf{u}) = \frac{1}{J}\sum_{j=1}^{J}\left|\sum_{k=0}^{K-1} A(\mathrm{R}^{k^T}\mathbf{u})\exp(i2\pi\mathbf{u}\cdot\mathbf{t}_{jk})\right|^2 , \tag{7.50}$$

and the ensemble averaged intensity at the ith iteration of the iterative phase retrieval algorithm subsequently becomes

$$I^{(i)}(\mathbf{u}) = \frac{1}{J}\sum_{j=1}^{J}\left|\sum_{k=0}^{K-1} A^{(i)}(\mathrm{R}^{k^T}\mathbf{u})\exp(i2\pi\mathbf{u}\cdot\mathbf{t}_{jk})\right|^2 . \tag{7.51}$$

The real space projection P_S involves applying a support and positivity constraint to $a(\mathbf{x})$. Real space symmetry constraints are now not needed since they are built in through Eq. (7.50). The Fourier space projection P_M is required to enforce the relationship given by Eq. (7.50). The similarity between the form of Eqs. (7.42) and (7.51) suggested to Kirian et al. [2014] that the appropriate Fourier space projection might be given by

$$A^{(i+1)}(\mathbf{u}) = \sqrt{\frac{I(\mathbf{u})}{I^{(i)}(\mathbf{u})}}A^{(i)}(\mathbf{u}) , \tag{7.52}$$

where $I^{(i)}(\mathbf{u})$ is calculated using Eq. (7.51).

The validity of Eq. (7.52) is now investigated. One requirement for this validity is that application of Eq. (7.52) to the iterate at the ith iteration should give an intensity $I^{(i+1)}(\mathbf{u})$ that is equal to the data $I(\mathbf{u})$. This requirement is investigated here.

The averaged diffracted intensity over a total of J types of unit cells given by Eq. (7.50) can be expanded out and written as

$$\begin{aligned} I(\mathbf{u}) &= \frac{1}{J}\sum_{j=1}^{J}\left(\sum_{k=0}^{K-1}\left|A(\mathrm{R}^{k^T}\mathbf{u})\right|^2 + \sum_{k_1=0}^{K-1}\sum_{\substack{k_2=0\\k_1\neq k_2}}^{K-1} A(\mathrm{R}^{k_1^T}\mathbf{u})A^*(R^{k_2^T}\mathbf{u})\exp\left(i2\pi\mathbf{u}\cdot(\mathbf{t}_{jk_1}-\mathbf{t}_{jk_2})\right)\right) \\ &= \sum_{k=0}^{K-1}\left|A(\mathrm{R}^{k^T}\mathbf{u})\right|^2 + \sum_{k_1=0}^{K-1}\sum_{\substack{k_2=0\\k_1\neq k_2}}^{K-1}\left(A(\mathrm{R}^{k_1^T}\mathbf{u})A^*(\mathrm{R}^{k_2^T}\mathbf{u})\frac{1}{J}\sum_{j=1}^{J}\exp\left(i2\pi\mathbf{u}\cdot(\mathbf{t}_{jk_1}-\mathbf{t}_{jk_2})\right)\right). \end{aligned} \tag{7.53}$$

Assume for the moment that,

$$I(\mathrm{R}^{k^T}\mathbf{u}) = I(\mathbf{u}) \tag{7.54}$$

for all $k \in \{0,\cdots,K-1\}$. The Fourier intensity of the $(i{+}1)$th iterate after the application of the projection P_M on the ith iterate is then

$$\begin{aligned} I^{(i+1)}(\mathbf{u}) &= \sum_{k=0}^{K-1}\left|A^{(i)}(\mathrm{R}^{k^T}\mathbf{u})\sqrt{\frac{I(\mathrm{R}^{k^T}\mathbf{u})}{I^{(i)}(\mathrm{R}^{k^T}\mathbf{u})}}\right|^2 + \sum_{k_1=0}^{K-1}\sum_{\substack{k_2=0\\k_1\neq k_2}}^{K-1}\left(A^{(i)}(\mathrm{R}^{k_1^T}\mathbf{u})\sqrt{\frac{I(\mathrm{R}^{k_1^T}\mathbf{u})}{I^{(i)}(\mathrm{R}^{k_1^T}\mathbf{u})}}\right. \\ &\quad \left.\times A^{(i)*}(\mathrm{R}^{k_2^T}\mathbf{u})\sqrt{\frac{I(\mathrm{R}^{k_2^T}\mathbf{u})}{I^{(i)}(\mathrm{R}^{k_2^T}\mathbf{u})}}\frac{1}{J}\sum_{j=1}^{J}\exp\left(i2\pi\mathbf{u}\cdot(\mathbf{t}_{jk_1}-\mathbf{t}_{jk_2})\right)\right). \\ &= \frac{I(\mathbf{u})}{I^{(i)}(\mathbf{u})}\sum_{k=0}^{K-1}\left|A^{(i)}(\mathrm{R}^{k^T}\mathbf{u})\right|^2 \\ &\quad + \frac{I(\mathbf{u})}{I^{(i)}(\mathbf{u})}\sum_{k_1=0}^{K-1}\sum_{\substack{k_2=0\\k_1\neq k_2}}^{K-1}\left(A^{(i)}(\mathrm{R}^{k_1^T}\mathbf{u})A^{(i)*}(\mathrm{R}^{k_2^T}\mathbf{u})\frac{1}{J}\sum_{j=1}^{J}\exp\left(i2\pi\mathbf{u}\cdot(\mathbf{t}_{jk_1}-\mathbf{t}_{jk_2})\right)\right). \\ &= \frac{I(\mathbf{u})}{I^{(i)}(\mathbf{u})}I^{(i)}(\mathbf{u}) \\ &= I(\mathbf{u}). \end{aligned} \tag{7.55}$$

Therefore, subject to Eq. (7.54), the modification due to the projection P_M given by Eq. (7.52) brings the diffracted intensity of the iterate to $I(\mathbf{u})$ as desired.

Looking now at Eq. (7.54), let α be an element of the set of integers $\{0, \cdots, K-1\}$. Then by Eq. (7.53),

$$\begin{aligned}
I(\mathrm{R}^{\alpha^T}\mathbf{u}) &= \sum_{k=0}^{K-1} \left|A(\mathrm{R}^{k^T}\mathrm{R}^{\alpha^T}\mathbf{u})\right|^2 \\
&\quad + \sum_{k_1=0}^{K-1}\sum_{\substack{k_2=0\\k_1\neq k_2}}^{K-1} \left(A(\mathrm{R}^{k_1^T}\mathrm{R}^{\alpha^T}\mathbf{u})A^*(\mathrm{R}^{k_2^T}\mathrm{R}^{\alpha^T}\mathbf{u})\frac{1}{J}\sum_{j=1}^{J}\exp\left(i2\pi(\mathrm{R}^{\alpha^T}\mathbf{u})\cdot(\mathbf{t}_{jk_1}-\mathbf{t}_{jk_2})\right)\right) \\
&= \sum_{k=0}^{K-1} \left|A(\mathrm{R}^{(\alpha+k)^T}\mathbf{u})\right|^2 \\
&\quad + \sum_{k_1=0}^{K-1}\sum_{\substack{k_2=0\\k_1\neq k_2}}^{K-1} \left(A(\mathrm{R}^{(\alpha+k_1)^T}\mathbf{u})A^*(\mathrm{R}^{(\alpha+k_2)^T}\mathbf{u})\frac{1}{J}\sum_{j=1}^{J}\exp\left(i2\pi(\mathrm{R}^{\alpha^T}\mathbf{u})\cdot(\mathbf{t}_{jk_1}-\mathbf{t}_{jk_2})\right)\right) \\
&= \sum_{k=\alpha}^{K-1+\alpha} \left|A(\mathrm{R}^{k^T}\mathbf{u})\right|^2 \\
&\quad + \sum_{k_1=\alpha}^{K-1+\alpha}\sum_{\substack{k_2=\alpha\\k_1\neq k_2}}^{K-1+\alpha} \left(A(\mathrm{R}^{k_1^T}\mathbf{u})A^*(\mathrm{R}^{k_2^T}\mathbf{u})\frac{1}{J}\sum_{j=1}^{J}\exp\left(i2\pi(\mathrm{R}^{\alpha^T}\mathbf{u})\cdot(\mathbf{t}_{j(k_1-\alpha)}-\mathbf{t}_{j(k_2-\alpha)})\right)\right).
\end{aligned} \tag{7.56}$$

The rotation/reflection operations for crystallographic unit cells form a group. Specifically, they are closed, and $\mathrm{R}^{K+\alpha} = \mathrm{R}^{\alpha}$, which means that

$$\sum_{k=\alpha}^{K-1+\alpha} = \sum_{k=\alpha}^{K-1} + \sum_{k=0}^{\alpha-1} = \sum_{k=0}^{K-1}. \tag{7.57}$$

Utilising Eq. (7.57) in Eq. (7.56) gives

$$\begin{aligned}
I(\mathrm{R}^{\alpha^T}\mathbf{u}) &= \sum_{k=0}^{K-1} \left|A(\mathrm{R}^{k^T}\mathbf{u})\right|^2 \\
&\quad + \sum_{k_1=0}^{K-1}\sum_{\substack{k_2=0\\k_1\neq k_2}}^{K-1} \left(A(\mathrm{R}^{k_1^T}\mathbf{u})A^*(\mathrm{R}^{k_2^T}\mathbf{u})\frac{1}{J}\sum_{j=1}^{J}\exp\left(i2\pi(\mathrm{R}^{\alpha^T}\mathbf{u})\cdot(\mathbf{t}_{j(k_1-\alpha)}-\mathbf{t}_{j(k_2-\alpha)})\right)\right)
\end{aligned} \tag{7.58}$$

which by comparison with Eq. (7.53) reduces to $I(\mathbf{u})$ if

$$\sum_{j=1}^{J}\exp\left(i2\pi(\mathrm{R}^{\alpha^T}\mathbf{u})\cdot(\mathbf{t}_{j(k_1-\alpha)}-\mathbf{t}_{j(k_2-\alpha)})\right) = \sum_{j=1}^{J}\exp\left(i2\pi\mathbf{u}\cdot(\mathbf{t}_{jk_1}-\mathbf{t}_{jk_2})\right). \tag{7.59}$$

Therefore, if Eq. (7.59) is satisfied, then Eq. (7.54) is also satisfied.

Rewriting the dot product in the argument in the exponential on the left-hand-side of Eq. (7.59) shows what is needed is that

$$\mathrm{R}^{\alpha}(\mathbf{t}_{j(k_1-\alpha)} - \mathbf{t}_{j(k_2-\alpha)}) = \mathbf{t}_{jk_1} - \mathbf{t}_{jk_2}, \tag{7.60}$$

for all α, k_1 and k_2 elements of the set of integers $\{0, \cdots, K-1\}$ with $k_1 \neq k_2$, and $k_i - \alpha$ is taken modulo K.

For the case of the P4 symmetry considered in Sec. 7.3.1, the set of $\mathbf{t}_{jk}$ can be shown to satisfy Eq. (7.60) by testing each of the J combinations. The difference vectors are shown in Fig. 7.22. For example for $\alpha = 1$, $k_1 = 2$ and $k_2 = 3$, $\mathbf{t}_{j(k_1-\alpha)} - \mathbf{t}_{j(k_2-\alpha)} = \mathbf{t}_{11} - \mathbf{t}_{12}$ and $\mathbf{t}_{12} - \mathbf{t}_{13}$, and they can be shown to be related by a rotation given by R^1 as seen in Fig. 7.22. Similar diagrams can be drawn for j=2,3,4. Inspection of the figure shows that Eq. (7.60) is satisfied.

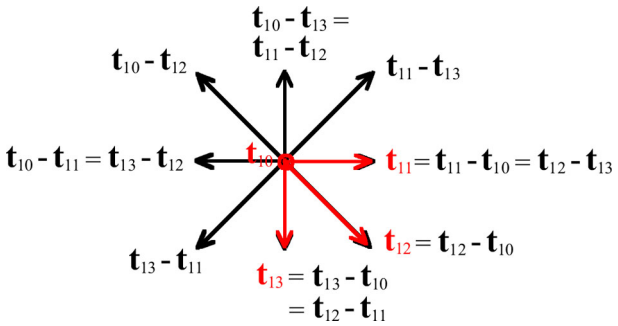

Figure 7.22 Difference vectors $\mathbf{t}_{jk_1} - \mathbf{t}_{jk_2}$ for the intuitive definition of the P4 symmetry translation vectors shown in Fig. 7.20 for $j = 1$.

These results provide some support for the validity of the projection given by Eq. (7.52), but considerably more work needs to be done to verify that this is the correct projection, specifically it may not be distance minimising and it may not be valid for all space groups.

7.4 CONCLUDING REMARKS

Retrieving the phases of multiple objects from their averaged intensity is possible. This kind of problem arises, for example, if diffraction intensities are measured from individual objects with low SNR and it is necessary to average the measurements over multiple objects in order to increase the SNR to a useable level. A similar situation could occur if there is a single kind of object, but it adopts one of a small number of distinct, unknown configurations. Summing over many different patterns, as is necessary in SFX, means that the resulting continuous diffraction is equal to the incoherent average over the set of all possible unit cells. All of these aforementioned problems are related in that they involve multiple-object phase retrieval.

In general, uniqueness of phase retrieval from the summed magnitude of multiple objects depends on the constraint ratio Ω which derives from the supports of the individual objects and the dimensionality of the problem. $\Omega > 1$ is needed for unique reconstructions. However uniqueness problems persist for objects that have identical supports even if $\Omega > 1$. In the case of all objects having identical supports, any linear combination of the original set of objects and their inversion about the origin such that the coefficient matrices satisfy Eqs. (7.17) and (7.18) have the same averaged diffracted intensity. Objects with identical but non-centrosymmetric supports lessens the non-uniqueness by allowing only half as much mixing to occur. Positivity also substantially reduces the non-uniqueness of the phase retrieval problem, in particular for objects with a wide range of sample values.

For problems where a unique solution is expected, there exists a straightforward implementation of iterative projection algorithms for phase retrieval that incorporates object support, positivity and the averaged Fourier intensity constraints. Simulations with no noise show that this algorithm is effective, and indeed the objects can be recovered in cases where a unique solution is expected based on the theory described.

The analogous problem in SFX is when phases are to be retrieved from an ensemble averaged and finely sampled intensity obtained from crystals with more than one molecules in the unit cell. This corresponds to phase retrieval for multiple objects where the objects are the different kinds of unit cell. The multiple objects in this case are related, both by the symmetry within, and the relationships between, the different kinds of unit cell.

There are two approaches to this type of phase retrieval problem. The first follows that described for multiple independent objects, and reconstructs all the different kinds of unit cell while enforcing the relationships within and between them in real space. This approach was shown to be effective and was able to indicate that the reconstruction problem has some non-uniqueness for the 2-D case investigated with P4 symmetry. The second approach is to use an iteration that reconstructs the single asymmetric unit using a reciprocal space projection that incorporates the real space relationships. The asymmetric unit is now the iterate in the IPA used for solving the problem. The first approach is conceptually simpler but requires more computational resources in terms of taking J times more Fourier transform operations for reconstructing J unit cells, and furthermore, over a larger array of numbers as the set of unit cells is iterated, rather than the asymmetric unit. A more detailed analysis of this method is required.

Note also that the unit cells generated from translationally wrapping the asymmetric unit of the plane P4 space group studied in this chapter are in fact all centrosymmetric, implying that their Fourier transform is real with phase values restricted to either 0 or π, rendering the phase problem easier to solve in general. More work is needed to study cases where the generated unit cells are noncentrosymmetric.

The ambiguous solutions that arise in phase retrieval from the averaged diffracted intensity of many different unit cells could be resolved in practice by taking into account information associated with the physical structures of real molecules, given reconstructions of high enough resolution.

Chapter 8

CONCLUSIONS AND SUGGESTIONS FOR FUTURE WORK

The recent development of the X-ray free-electron laser has the potential to solve three of the major problems of conventional X-ray crystallography; that of (1) crystallisation difficulties, (2) radiation damage, and (3) the phase problem, via its ability to generate intense but extremely brief pulses of X-rays. The high intensity X-ray pulses enable measurable diffraction data to be obtained from crystals only a few unit cells across. Such small crystals have the advantage that they can aid in the solution of the phase problem by providing information between the Bragg peaks. The method of serial femtosecond nanocrystallography that has arisen from these developments has prompted the exploration of novel ideas to process and interpret the data resulting from these experiments.

Serial femtosecond nanocrystallography using XFELs requires individual diffraction patterns from nanocrystals to be averaged and merged into a 3-D volume. The resulting data is the diffracted intensity from the unit cell modulated by the averaged interference function, also known as the averaged shape transform, a function that describes the mean influence from the varying sizes and shapes of the nanocrystals on their diffraction. The data can potentially be processed to estimate the magnitude of the Fourier transform of one unit cell, offering the possibility of direct phasing of the diffraction data. The estimate of the unit cell transform obtained in this way is noise-sensitive however at positions between the Bragg reflections where the value of the averaged shape transform is small. Using samples of the processed data that have a higher signal-to-noise ratio while retaining enough samples to keep the phase retrieval problem well-determined, allow the effects of noise to be reduced.

Nanocrystals in non-trivial space groups will generally form crystals with unit cells that have different contents, and with incomplete unit cells on the crystals' surface. The diffraction from an ensemble of such crystals depends on the average over the different possible unit cells and contains extra components that depend on the incomplete unit cell surface structure. Theory describing these diffraction components has been developed and numerical studies indicate that effects of the incomplete unit cells on

the crystal surface is small and that the diffraction is dominated by the average over the different kinds of complete unit cells that are present.

The problem of reconstructing a set of objects from their averaged Fourier intensities is directly related to that of reconstructing multiple unit cells. Despite the loss of information as a result of the averaging, there are many cases where a unique solution is to be expected. The uniqueness properties of this problem have been characterised in detail and depends primarily on the ratio of the number of independent data to the number of independent object parameters, characterised by the constraint ratio Ω, which is a function of the object supports and the dimensionality. A constraint ratio of $\Omega > 1$ is necessary for a unique solution and is a sufficient condition if the object supports are different. It is not sufficient however if any of the object supports are the same, since appropriate linear combinations of the objects with the same support also represent solutions that are consistent with the averaged intensity. A reconstruction algorithm has been described and simulations show that it is effective. The problem of phase retrieval in the presence of multiple unit cells is similar to that for multiple independent objects, although has some different characteristics. Some of these characteristics have been explored but more work is needed.

8.1 SUGGESTIONS FOR FURTHER RESEARCH

1. Phase retrieval for multiple unit cells. Fundamental questions such as how many different kinds of unit cells can be reconstructed given their summed Fourier intensity and symmetry relationships need to be answered. Furthermore, the case studied in this thesis happens to be centrosymmetric, therefore more work needs to be done to explore cases that are noncentrosymmetric.

2. A rigorous analysis of the projection algorithm proposed by Kirian et al. [2014] is needed.

3. An algorithm for finding the optimal unit cell that minimises the integrated electron density of incomplete unit cells on the surface of a crystal is of interest.

4. Importantly, the kinds of methods discussed in this thesis need to be attempted and evaluated with experimental data.

REFERENCES

Abdallah, B.G., Chao, T.C., Kupitz, C., Fromme, P. and Ros, A. (2013), 'Dielectrophoretic sorting of membrane protein nanocrystals', *ACS Nano*, Vol. 7, No. 10, pp. 9129–9137.

Ackermann, W., Asova, G., Ayvazyan, V. et al. (2007), 'Operation of a free-electron laser from the extreme ultraviolet to the water window', *Nature Photonics*, Vol. 1, pp. 336–342.

Adams, W.J. (1974), *The Life and Times of the Central Limit Theorem*, Kaedmon Pub. Co., New York, USA.

Aherne, M., Lyons, J.A. and Caffrey, M. (2012), 'A fast, simple and robust protocol for growing crystals in the lipidic cubic phase', *J. Appl. Cryst.*, Vol. 45, pp. 1330–1333.

Als-Nielsen, J. and McMorrow, D. (2011), *Elements of Modern X-ray Physics*, John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester, West Sussex, UK, 2nd ed.

Altarelli, M., Brinkmann, R., Chergui, M. et al. (2006), 'XFEL: The European X-ray Free-electron Laser - Technical Design Report', *DESY Report*, No. 2006-097, July. Available online at http://www.xfel.eu/documents/ Retrieved: 31 Oct 2012.

Barakat, R. and Newsam, G. (1984), 'Necessary conditions for a unique solution to two-dimensional phase recovery', *J. Math. Phys.*, Vol. 25, No. 11, pp. 3190–3193.

Barends, T.R.M., Foucar, L., Botha, S. et al. (2013), 'De novo protein crystal structure determination from X-ray free-electron laser data', *Nature*, Vol. 505, pp. 244–247.

Barton, D. (2002), *Year 13 Statistics*, Longman Write-on Notes, Pearson Education New Zealand Limited, 46 Hillside Road, Auckland, NZ, 3rd ed.

Barty, A., Caleman, C., Chapman, H.N. et al. (2012), 'Self-terminating diffraction gates femtosecond X-ray nanocrystallography measurements', *Nature Photonics*, Vol. 6, pp. 35–40.

Barty, A., Chapman, H.N., White, T.A. et al. (2013), 'dataXpress: Processing the deluge of data from XFEL', *2013 APS/CNM/EMC Users Meeting Workshop*.

Barty, A., Kirian, R.A., Maia, F.R.N.C., Hantke, M., Yoon, C.H., White, T.A. and Chapman, H.N. (2014), 'Cheetah: Software for high-throughput reduction and analysis of serial femtosecond X-ray diffraction data', *J. Appl. Cryst.*, Vol. 47, pp. 1118–1131.

Bates, R.H.T. (1982), 'Fourier phase problems are uniquely solvable in more than one dimension I. Underlying theory', *Optik*, Vol. 61, No. 3, pp. 247–262.

Blahut, R.E. (2011), *Theory of Remote Image Formation*, Cambridge University Press, The Edinburgh Building, Cambridge, UK, paperback ed.

Blewett, J. (1998), 'Synchrotron radiation - Early history', *J. Synchrotron Rad.*, Vol. 5, pp. 135–139.

Bogan, J., Boutet, S., Chapman, H.N. et al. (2010), 'Aerosol imaging with a soft x-ray free-electron laser', *Aerosol Science and Technology*, Vol. 44, No. 3, pp. I–VI.

Born, M. and Wolf, E. (2003), *Principles of Optics*, Cambridge University Press, Cambridge, UK, 7th ed.

Bortel, G. and Tegze, M. (2011), 'Common arc method for diffraction pattern orientation', *Acta Cryst.*, Vol. 67, pp. 533–543.

Boutet, S., Lomb, L., Williams, G. et al. (2012), 'High-resolution protein structure determination by serial femtosecond crystallography', *Science*, Vol. 337, pp. 362–364.

Bruck, Y. and Sodin, L. (1979), 'On the ambiguity of the image reconstruction problem', *Optics Communications*, Vol. 30, No. 3, pp. 304–308.

Bucksbaum, P.H. and Merlin, R. (1999), 'The phonon Bragg switch: a proposal to generate sub-picosecond X-ray pulses', *Solid State Communications*, Vol. 111, pp. 535–539.

Bullen, P.S. (2003), *Handbook of Means and Their Inequalities*, Kluwer Academic Publishers, 3300 AA Dordrecht, The Netherlands.

Caffrey, M., Li, D., Howe, N. and Shah, S.T.A. (2014), ''Hit and run' serial femtosecond crystallography of a membrane kinase in the lipid cubic phase', *Phil. Trans. R. Soc. B*, Vol. 369, No. 20130621.

Caleman, C., Huldt, G., Maia, F.R.N.C. et al. (2011), 'On the feasibility of nanocrystal imaging using intense and ultrashort X-ray pulses', *ACS. nano.*, Vol. 5, No. 1, pp. 139–146.

Carini, G.A., Boutet, S., Chollet, M. et al. (2014), 'Experience with the CSPAD during dedicated detector runs at LCLS', *17th Pan-American Synchrotron Radiation Instrumentation Conference, Journal of Physics: Conference Series*, Vol. 493, No. 012011.

Chapman, H.N. (2009), 'X-ray imaging beyond the limits', *Nature Materials*, Vol. 8, pp. 299–301.

Chapman, H.N., Barty, A., Bogan, M. et al. (2006a), 'Femtosecond diffractive imaging with a soft-x-ray free-electron laser', *Nature Physics*, Vol. 2, pp. 839–843.

Chapman, H.N., Barty, A., Marchesini, S. et al. (2006b), 'High-resolution ab initio three-dimensional x-ray diffraction microscopy', *J. Opt. Soc. Am. A*, Vol. 23, No. 5, pp. 1179–1200.

Chapman, H.N., Fromme, P., Barty, A. et al. (2011), 'Femtosecond X-ray protein nanocrystallography', *Nature*, Vol. 470, pp. 73–78.

Chapman, H.N., Caleman, C. and Timneanu, N. (2014), 'Diffraction before destruction', *Phil. Trans. R. Soc. B*, Vol. 369, No. 20130313.

Chen, J.P.J. and Millane, R.P. (2013), 'Diffraction by nanocrystals', *J. Opt. Soc. Am. A*, Vol. 30, No. 12, pp. 2627–2634.

Cusack, S., Belrhali, H., Bram, A., Burghammer, M., Perrakis, A. and Riekel, C. (1998), 'Small is beautiful - Protein micro-crystallography', *Nat. Struct. Biol.*, Vol. 5, pp. 634–637.

Deacon, D., Elias, L., Madey, J., Ramian, G., Schwettman, H. and Smith, T. (1977), 'First operation of a free-electron laser', *Phys. Rev. Lett.*, Vol. 38, No. 16, pp. 892–894.

DeLucas, L. (2009), *Membrane Protein Crystallization*, Academic Press, Burlington MA. USA.

Denes, P. (2014), 'Two-dimensional imaging detectors for structural biology with X-ray lasers', *Phil. Trans. R. Soc. B*, Vol. 369, No. 20130334.

DePonte, D.P., Weierstall, U., Schmidt, K. et al. (2008), 'Gas dynamic virtual nozzle for generation of microscopic droplet streams', *J. Phys. D: Appl. Phys.*, Vol. 41, No. 195505.

Dilanian, R.A., Streltsov, V.A., Quiney, H.M. and Nugent, K.A. (2013), 'Continuous X-ray diffractive field in protein nanocrystallography', *Acta Cryst. A*, Vol. 69, pp. 108–118.

Dodson, E. (2008), 'The befores and afters of molecular replacement', *Acta Cryst. D*, Vol. 64, pp. 17–24.

Dragone, A., Caragiulo, P., Markovic, B. et al. (2014), 'ePix: a class of architectures for second generation LCLS cameras', *17th Pan-American Synchrotron Radiation Instrumentation Conference, Journal of Physics: Conference Series*, Vol. 493, p. 012012.

Dreele, R.V. (2007), 'Multipattern rietveld refinement of protein powder data - An approach to higher resolution', *J. Appl. Cryst.*, Vol. 40, pp. 133–143.

Drenth, J. (2007), *Principles of Protein X-Ray Crystallography*, Springer, 233 Spring Street, New York, NY, USA, 3rd ed.

Eden, M. (1961), 'A two-dimensional growth process', *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 4: Biology and Problems of Health, pp. 223–239. University of California Press, Berkeley, Ed. J. Neyman.

Ekeberg, T., Svenda, M., Abergel, C. et al. (2015), 'Three-dimensional reconstruction of the giant mimivirus particle with an x-ray free-electron laser', *Phys. Rev. Lett.*, Vol. 114, p. 098102.

Elias, L., Fairbank, W., Madey, J. et al. (1976), 'Observation of stimulated emission of radiation by relativistic electrons in a spatially periodic transverse magnetic field', *Phys. Rev. Lett.*, Vol. 36, pp. 717–720.

Elser, V. (2003a), 'Random projections and the optimization of an algorithm for phase retrieval', *J. Phys. A-Math Gen.*, Vol. 36, pp. 2995–3007.

Elser, V. (2003b), 'Phase retrieval by iterated projections', *J. Opt. Soc. Am. A*, Vol. 20, No. 1, pp. 40–55.

Elser, V. (2013), 'Direct phasing of nanocrystal diffraction', *Acta Cryst. A*, Vol. 69, pp. 559–569.

Elser, V. and Millane, R.P. (2008), 'Reconstruction of an object from its symmetry-averaged diffraction pattern', *Acta Cryst. A*, Vol. 64, pp. 273–279.

Elser, V., Rankenburg, I. and Thibault, P. (2007), 'Searching with iterated maps', *PNAS*, Vol. 104, No. 2, pp. 418–423.

Emma, P., Akre, R., Arthur, J. et al. (2010), 'First lasing and operation of an ångstrom-wavelength free-electron laser', *Nature Photonics*, Vol. 4, Sep, pp. 641–647.

Ewald, P.P. (1969), 'Introduction to the dynamical theory of x-ray diffraction', *Acta Cryst. A*, Vol. 25, pp. 103–108.

Fienup, J.R. (1982), 'Phase retrieval algorithms: a comparison', *Appl. Opt.*, Vol. 21, No. 15, pp. 2758–2769.

Friedrich, W., Knipping, P. and von Laue, M. (1912), 'Interferenz-Erscheinungen bei Röntgenstrahlen', *Sitzungsberichte der Mathematisch-Physikalischen Classe der Königlich-Bayerischen Akademie der Wissenschaften zu München*, Vol. 92, pp. 303–322.

Fung, R., Shneerson, V., Saldin, D. and Ourmazd, A. (2009), 'Structure from fleeting illumination of faint spinning objects in flight', *Nature Physics*, Vol. 5, pp. 64–67.

Garman, E. (1999), 'Cool data - quantity and quality', *Acta Cryst. D*, Vol. 55, pp. 1641–1653.

Geloni, G., Kocharyan, V. and Saldin, E. (2011), 'A novel self-seeding scheme for hard X-ray FELs', *J. Modern Optics*, Vol. 58, No. 16, pp. 1391–1403.

Georgieva, D., Kuil, M., Oosterkamp, T., Zandbergen, H. and Abrahams, J. (2007), 'Heterogeneous nucleation of three-dimensional protein nanocrystals', *Acta Cryst. D*, Vol. 63, pp. 564–570.

Giacovazzo, C. (1999), *Direct phasing in Crystallography: Fundamentals and Applications*, Vol. 8 of International Union of Crystallography Monographs on Crystallography, Oxford University Press, USA.

Goodman, J.W. (1985), *Statistical Optics*, John Wiley & Sons, Inc., The Atrium, Southern Gate, Chichester, West Sussex, U.K.

Gravel, S. and Elser, V. (2008), 'Divide and concur: A general approach to constraint satisfaction', *Phys. Rev. E*, Vol. 78, p. 036706.

Green, D., Ingram, V. and Perutz, M. (1954), 'The structure of haemoglobin IV. Sign determination by the isomorphous replacement method', *Proc. R. Soc. London Ser. A*, Vol. 255, pp. 287–307.

Griffiths, D. (1981), *Introduction to Electrodynamics*, Prentice-Hall, Englewood Cliffs, NJ, USA, 3rd ed.

Hahn, T. (2005), *International Tables for Crystallography*, Vol. A: Space-group symmetry, Springer, 3300 AA Dordrecht, The Netherlands, 5th ed.

Harker, D. and Kasper, J. (1948), 'Phases of Fourier coefficients directly from crystal diffraction data', *Acta Cryst.*, Vol. 1, pp. 70–75.

Hart, P., Boutet, S., Carini, G. et al. (2012), 'The Cornell-SLAC pixel array detector at LCLS', *Nuclear Science Symposium, Medical Imaging Conference*, No. SLAC-PUB-15284.

Hauptman, H.A. (1991), 'The phase problem of x-ray crystallography', *Rep. Prog. Phys.*, Vol. 54, pp. 1427–1454.

He, B. (2009), *Two-dimensional X-ray Diffraction*, John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester, West Sussex, UK.

Henderson, R. (2006), 'The potential and limitations of neutrons, electrons and X-rays for atomic resolution microscopy of unstained biological molecules', *Quart. Rev. Biophys.*, Vol. 28, No. 2, pp. 171–193.

Herrmann, S., Boutet, S., Duda, B. et al. (2013a), 'CSPAD-140k: A versatile detector for LCLS experiments', *Nuc. Inst. Meth. Phys. Res. A*, Vol. 718, pp. 550–553.

Herrmann, S., Blaj, G., Carini, G. et al. (2013b), 'CSPAD upgrades at LCLS', *SLAC-PUB*, No. 15843.

Hunter, M.S., Segelke, B., Messerschmidt, M. et al. (2014), 'Fixed-target protein serial microcrystallography with an x-ray free electron laser', *Nature Scientific Reports*, Vol. 4, p. 6026.

Jackson, J. (1999), *Classical Electrodynamics*, John Wiley & Sons, Hoboken, NJ, USA, 3rd ed.

James, R.W. (1948), *The Optical Principles of the Diffraction of X-rays*, Vol. II, G. Bell and Sons Ltd, London.

Johansson, L.C., Arnlund, D., White, T.A. et al. (2012), 'Lipidic phase membrane protein serial femtosecond crystallography', *Nature Methods*, Vol. 9, pp. 263–265.

Johansson, L.C., Arnlund, D., Katona, G. et al. (2013), 'Structure of a photosynthetic reaction centre determined by serial femtosecond crystallography', *Nature Communications*, Vol. 4, No. 2911.

Jordan, P., Fromme, P. and Witt, H.T. (2001), 'Three-dimensional structure of cyanobacterial photosystem I at 2.5Å resolution', *Nature*, Vol. 411, pp. 909–917.

Jurek, Z., Faigel, G. and Tegze, M. (2004), 'Dynamics in a cluster under the influence of intense femtosecond hard x-ray pulses', *Eur. Phys. J. D*, Vol. 29, pp. 217–229.

Kabsch, W. (1993), 'Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants', *J. Appl. Cryst.*, Vol. 26, pp. 795–800.

Kato, M., Tanaka, T., Kurosawa, T. et al. (2012), 'Pulse energy measurement at the hard x-ray laser in Japan', *Appl. Phys. Lett.*, Vol. 101, p. 023503.

Kerna, J., Alonso-Morib, R., Hellmichc, J. et al. (2012), 'Room temperature femtosecond X-ray diffraction of photosystem II microcrystals', *Proc. Natl Acad. Sci. USA*, Vol. 109, pp. 9721–9726.

Kirian, R.A. (2011), *Femtosecond X-ray Protein Nanocrystallography & Correlated Fluctuation Small-Angle X-ray Scattering*, PhD thesis, Arizona State University.

Kirian, R.A., Wang, X., Weierstall, U. et al. (2010), 'Femtosecond protein nanocrystallography - data analysis methods', *Optics Express*, Vol. 18, No. 6, pp. 5713–5723.

Kirian, R.A., White, T.A., Holton, J. et al. (2011), 'Structure-factor analysis of femtosecond micro-diffraction patterns from protein nanocrystals', *Acta Cryst. A*, Vol. 67, pp. 131–140.

Kirian, R.A., Bean, R.J., Beyerlein, K.R., Yefanov, O.M., White, T.A., Barty, A. and Chapman, H.N. (2014), 'Phasing coherently illuminated nanocrystals bounded by partial unit cells', *Phil. Trans. R. Soc. B*, Vol. 369, No. 20130331.

Kirian, R.A., Bean, R.J., Beyerlein, K.R. et al. (2015), 'Direct phasing of finite crystals illuminated with a free-electron laser', *Phys. Rev. X*, Vol. 5, p. 011015.

Koch, A., Kuster, M., Sztuk-Dambietz, J. and Turcato, M. (2013), 'Detector development for the European XFEL: Requirements and status', *11th International Conference on Synchrotron Radiation Instrumentation, Journal of Physics: Conference Series*, Vol. 425, No. 062013.

Landau, E.M. and Rosenbusch, J.P. (1996), 'Lipidic cubic phases: A novel concept for the crystallization of membrane proteins', *Proc. Natl. Acad. Sci. USA*, Vol. 93, pp. 14532–14535.

Lattman, E.E. and Loll, P.J. (2008), *Protein Crystallography - A Concise Guide*, John Hopkins University Press, 2715 North Charles st, Baltimore, Maryland, USA.

Lavelle, C. (2009), 'Left is right, right is wrong', *EMBO Reports*, Vol. 10, No. 11, pp. 1185–1186.

Lehninger, A.L., Nelson, D.L. and Cox, M.M. (2013), *Principles of Biochemistry*, W. H. Freeman, New York, N.Y., USA, 6th ed.

Leslie, A. (1991), 'Molecular Data Processing', In Moras, D., Podjarny, A. and Thierry, J. (Eds.), *Crystallographic computing 5 - From chemistry to biology: papers presented at the International School on Crystallographics Computing held at Bischneberg, France, 29 July-5 August, 1990*, Oxford University Press, p. 50.

Leslie, A. (2006), 'The integration of macromolecular diffraction data', *Acta Cryst.*, Vol. 62, pp. 48–57.

Limpert, E., Stahel, W.A. and Abbt, M. (2001), 'Log-normal distributions across the sciences: Keys and clues', *BioScience*, Vol. 51, pp. 341–352.

Littke, W. and John, C. (1984), 'Protein single crystal growth under microgravity', *Science*, Vol. 225, No. 4658, pp. 203–204.

Liu, W., Wacker, D., Gati, C. et al. (2013), 'Serial femtosecond crystallography of G protein-coupled receptors', *Science*, Vol. 342, pp. 1521–1524.

Liu, H., Zatsepin, N.A. and Spence, J.C.H. (2014), 'Ab-initio phasing using nanocrystal shape transforms with incomplete unit cells', *IUCrJ*, Vol. 1, pp. 19–27.

Loh, N.T.D. and Elser, V. (2009), 'Reconstruction algorithm for single-particle diffraction imaging experiments', *Phys. Rev. E*, Vol. 80, p. 026705.

Lomb, L., Steinbrener, J., Bari, S. et al. (2012), 'An anti-settling sample delivery instrument for serial femtosecond crystallography', *J. Appl. Cryst.*, Vol. 45, pp. 674–678.

Madey, J. (1971), 'Stimulated emission of bremsstrahlung in a periodic magnetic field', *J. Appl. Phys.*, Vol. 42, pp. 1906–1913.

Maia, F.R.N.C. (2012), 'The Coherent X-ray Imaging Data Bank', *Nature Methods*, Vol. 9, pp. 854–855.

Marchesini, S., He, H. and Chapman, H.N. (2003), 'X-ray image reconstruction from a diffraction pattern alone', *Phys. Rev. B*, Vol. 68, p. 140101.

Margaritondo, G. and Ribic, P.R. (2011), 'A simplified description of X-ray free-electron lasers', *J. Synchrotron Rad.*, Vol. 18, pp. 101–108.

Martin, A.V., Wang, F., Loh, N.D. et al. (2012), 'Noise-robust coherent diffractive imaging with a single diffraction pattern', *Optics Express*, Vol. 20, No. 15, pp. 16650–16661.

McNeil, B.W.J. and Thompson, N.R. (2010), 'X-ray free-electron lasers', *Nature Photonics*, Vol. 4, Dec, pp. 814–821.

Miao, J., Charalambous, P., Kirz, J. and Sayre, D. (1999), 'Extending the methodology of X-ray crystallography to allow imaging of micrometre-sized non-crystalline specimens', *Nature*, Vol. 400, pp. 324–344.

Millane, R.P. (1990), 'Phase retrieval in crystallography and optics', *J. Opt. Soc. Am. A*, Vol. 7, No. 3, pp. 394–411.

Millane, R.P. and Hsiao, W. (2009), 'The basis of phase dominance', *Optics letters*, Vol. 34, No. 17, pp. 2607–2609.

Millane, R.P. and Lo, V.L. (2013), 'Iterative projection algorithms in protein crystallography. I. Theory', *Acta Cryst.*, Vol. 69, pp. 517–527.

Millane, R.P. and Stroud, W.J. (1997), 'Reconstructing symmetric images from their undersampled Fourier intensities', *J. Opt. Soc. Am. A*, Vol. 14, No. 3, pp. 568–579.

Motz, H., Thon, W. and Whitehurst, R. (1953), 'Experiments on radiation by fast electron beams', *J. Appl. Phys.*, Vol. 24, pp. 826–833.

Neutze, R., Wouts, R., van der Spoel, D., Weckert, E. and Hajdu, J. (2000), 'Potential for biomolecular imaging with femtosecond x-ray pulses', *Nature*, Vol. 406, pp. 752–757.

Otwinowski, Z. and Minor, W. (1997), 'Processing of X-ray diffraction data collected in oscillation mode', *Methods Enzymol*, Vol. 276, pp. 307–326.

Pflugrath, J. (1999), 'The finer things in X-ray diffraction data collection', *Acta Cryst.*, Vol. 55, pp. 1718–1725.

Philipp, H., Koerner, L., Hromalik, M. et al. (2010), 'Femtosecond radiation experiment detector for X-ray free-electron laser (XFEL) coherent X-ray imaging', *IEEE Trans. Nucl. Sci.*, Vol. 57, No. 6, pp. 3795–3799.

Plancherel, M. (1910), 'Contribution à l'étude de la représentation d'une fonction arbitraire par des intégrales définies', *Rend. Circ. Matem. Palermo*, Vol. 30, pp. 289–335.

Ravelli, R. and Garman, E. (2006), 'Radiation damage in macromolecular cryocrystallography', *Curr. Opin. Struc. Bio.*, Vol. 16, pp. 624–629.

Rayleigh, L. (1879), 'On the capillary phenomena of jets', *Proc. R. Soc. Lond.*, Vol. 29, pp. 71–97.

Redecke, L., Nass, K., DePonte, D.P. et al. (2013), 'Natively inhibited Trypanosoma brucei cathepsin B structure determined by using an X-ray laser', *Science*, Vol. 339, pp. 227–230.

Ren, G., Cheng, A., Reddy, V., Melnyk, P. and Mitra, A.K. (2000), 'Three-dimensional fold of the human AQP1 water channel determined at 4Å resolution by electron crystallography of two-dimensional crystals embedded in ice', *J. Mol. Biol.*, Vol. 301, pp. 369–387.

Ribic, P.R. and Margaritondo, G. (2012), 'Status and prospects of x-ray free-electron lasers (X-FELs): a simple presentation', *J. Phys. D: Appl. Phys.*, Vol. 45, p. 213001 (20pp).

Rossmann, M. and van Beek, C. (1999), 'Data processing', *Acta Cryst.*, Vol. 55, pp. 1631–1640.

Saldin, E.L., Schneidmiller, E.A. and Yurkov, M.V. (2000), *The Physics of Free-electron Lasers*, Springer-Verlag, Heidelberg, Germany.

Saldin, E.L., Schneidmiller, E.A. and Yurkov, M.V. (2004), 'Terawatt-scale sub-10-fs laser technology: Key to generation of GW-level attosecond pulses in X-ray free electron laser', *Optics Communications*, Vol. 237, pp. 153–164.

Sauter, N.K., Hattne, J., Grosse-Kunstleve, R.W. and Echolsa, N. (2013), 'New Python-based methods for data processing', *Acta Cryst. D*, Vol. 69, pp. 1274–1282.

Sayre, D. (1952), 'Some implications of a theorem due to Shannon', *Acta Cryst.*, Vol. 5, p. 843.

Sayre, D., Chapman, H.N. and Miao, J. (1998), 'On the extendibility of X-ray crystallography to noncrystals', *Acta Cryst. A*, Vol. 54, pp. 232–239.

Schoenlein, R., Chattopadhyay, S., Chong, H., Glover, T., Heimann, P., Shank, C., Zholents, A. and Zolotorev, M. (2000), 'Generation of femtosecond pulses of synchrotron radiation', *Science*, Vol. 287, pp. 2237–2240.

Seibert, M., Ekeberg, T., Maia, F.R.N.C. et al. (2011), 'Single mimivirus particles intercepted and imaged with an x-ray laser', *Nature*, Vol. 470, pp. 78–81.

Shannon, C.E. (1949), 'Communication in the presence of noise', *Proc. Inst. Radio Engrs.*, Vol. 37, No. 1, pp. 10–21.

Sherwood, D. and Cooper, J. (2011), *Crystals, X-rays and Proteins*, Oxford University Press, Great Clarendon Street, Oxford, UK.

Sierra, R.G., Laksmono, H., Kern, J. et al. (2012), 'Nanoflow electrospinning serial femtosecond crystallography', *Acta Cryst. D*, Vol. 68, pp. 1584–1587.

Snell, E., Weisgerber, S., Helliwell, J. et al. (1995), 'Improvements in lysozyme protein crystal perfection through microgravity growth', *Acta Cryst. D*, Vol. 51, pp. 1099–1102.

Solem, J. (1986), 'Imaging biological specimens with high-intensity soft x-rays', *Science*, Vol. 3, No. 11, pp. 1551–1565.

Son, S., Chapman, H.N. and Santra, R. (2011), 'Multiwavelength anomalous diffraction at high x-ray intensity', *Phys. Rev. Lett.*, Vol. 107, No. 21, p. 218102.

Son, S., Chapman, H.N. and Santra, R. (2013), 'Determination of multiwavelength anomalous diffraction coefficients at high x-ray intensity', *J. Phys. B: At. Mol. Opt. Phys.*, Vol. 46, p. 164015.

Spence, J.C.H. (2008), *Science of microscopy*, Springer, 233 Spring Street, New York, NY, USA, Chap. Diffractive (Lensless) Imaging.

Spence, J.C.H. and Doak, R.B. (2004), 'Single molecule diffraction', *Phys. Rev. Lett.*, Vol. 92, No. 19, p. 198102.

Spence, J.C.H., Kirian, R.A., Wang, X. et al. (2011), 'Phasing of coherent femtosecond x-ray diffraction from size-varying nanocrystals', *Optics Express*, Vol. 19, No. 4, pp. 2866–2873.

Spence, J.C.H., Weierstall, U. and Chapman, H.N. (2012), 'X-ray lasers for structural and dynamic biology', *Rep. Prog. Phys.*, Vol. 75, p. 102601.

Steller, I., Bolotovsky, R. and Rossmann, M. (1997), 'An algorithm for automatic indexing of oscillation images using Fourier analysis', *J. Appl. Cryst.*, Vol. 30, pp. 1036–1040.

Struder, L., Eppa, S., Rolles, D. et al. (2010), 'Large-format, high speed, x-ray pnCCDs combined with electron and ion imaging spectrometers in a multipurpose chamber for experiments at 4th generation light sources', *Nuc. Instr. Meth. Phys. Res. A*, Vol. 614, No. 3, pp. 483–496.

Susskind, L. (2005), *The Cosmic Landscape: String Theory and the Illusion of Intelligent Design*, Little, Brown and Co., New York, USA.

Thibault, P. (2007), *Algorithmic Methods in Diffraction Microscopy*, PhD thesis, Cornell University.

Thibault, P. and Elser, V. (2010), 'X-ray diffraction microscopy', *Annu. Rev. Condens. Matter Phys.*, Vol. 1, pp. 237–255.

Uson, I. and Sheldrick, G.M. (1999), 'Advances in direct methods for protein crystallography', *Curr. Opin. Struct. Biol.*, Vol. 9, pp. 643–648.

von Neumann, J. (1950), *Functional Operators*, Vol. II: The Geometry of Orthogonal Spaces of Ann. Math. Stud., Princeton University Press, Princeton, NJ, USA. Reprint of mimeographed lecture notes first distributed in 1933.

Wade, D. (2006), *Symmetry: The Ordering Principle*, Wooden Books, Walker & Co., USA.

Weierstall, U., Doak, R.B., Spence, J.C.H. et al. (2008), 'Droplet streams for serial crystallography of proteins', *Experiments in Fluids*, Vol. 44, No. 5, pp. 675–689.

Weierstall, U., Spence, J.C.H. and Doak, R.B. (2012), 'Injector for scattering measurements on fully solvated biospecies', *Rev. Sci. Inst.*, Vol. 83, p. 035108.

Weierstall, U., James, D., Wang, C. et al. (2014), 'Lipidic cubic phase injector facilitates membrane protein serial femtosecond crystallography', *Nature Communications*, Vol. 5, p. 3309.

Welberry, T.R. (2004), *Diffuse X-ray Scattering and Models of Disorder*, Oxford University Press, Great Clarendon Street, Oxford, UK.

Wheelon, A. (1968), *Tables of Summable Series and Integrals Involving Bessel Functions Part I*, Holden-Day Advanced Physics Monographs, Holden-Day, San Francisco, USA.

White, T.A. (2014), 'Post-refinement method for snapshot serial crystallography', *Phil. Trans. R. Soc. B*, Vol. 369, No. 20130330.

White, T.A., Kirian, R.A., Martin, A.V. et al. (2012), 'CrystFEL: a software suite for snapshot serial crystallography', *J. Appl. Cryst.*, Vol. 45, pp. 335–341.

White, T.A., Barty, A., Stellato, F. et al. (2013), 'Crystallographic data processing for free-electron laser sources', *Acta Cryst. D*, Vol. 69, pp. 1231–1240.

Yefanov, O.M., Gati, C., Bourenkov, G., Kirian, R.A., White, T.A., Spence, J.C.H., Chapman, H.N. and Barty, A. (2014), 'Mapping the continuous reciprocal space intensity distribution of X-ray serial crystallography', *Phil. Trans. R. Soc. B*, Vol. 369, No. 20130333.

Zayed, A. (1993), *Advances in Shannon's Sampling Theory*, CRC Press, Boca Raton, Florida, USA.